WITH REFERENCE TO TRUTH:

STUDIES IN REFERENTIAL SEMANTICS

by

DOUGLAS FILLMORE CANNON

A.B., Harvard University
1973

SUBMITTED TO THE DEPARTMENT OF
LINGUISTICS AND PHILOSOPHY
IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1982

Signature of Author ____________________
Department of Linguistics and Philosophy
April 8, 1982

Certified by ____________________
George Boolos
Thesis Supervisor

Accepted by ____________________
Judith Jarvis Thomson
Chairman, Department Graduate Committee

WITH REFERENCE TO TRUTH:

STUDIES IN REFERENTIAL SEMANTICS

by

DOUGLAS FILLMORE CANNON

Submitted to the Department of Linguistics and Philosophy on April 8, 1982, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Philosophy

ABSTRACT

In the first parts of my thesis I explore two philosophical programs in the area of referential semantics, namely, rigid designation accounts of proper names and naturalistic theories of truth. I conclude with an inquiry into the theory of truth for mathematics and its relationship to mathematical Platonism.

In Part One, I confront Kripke's well-known views with Quine's proposal that proper names correspond to a kind of predicate. I argue that the belief that proper names are rigid designators is unjustified and that many questions about the reference of terms in various possible worlds have no determinate answer. I take issue with Kripke's emphasis on the question, "How is the reference of names determined?", and suggest that it reflects dubious philosophical presuppositions.

In Part Two, I note that Field envisions a theory of truth as a natural property to be employed in a scientific account of the nature of knowledge. I explain how his account reworks the machinery of Tarski's theory of truth, making use of putative semantic relations between names and their referents and predicates and their extensions. It thereby commits itself to a natural relation of reference that underlies all language. I argue that nothing in our ordinary notion of referring provides evidence for such a relation and object to specific aspects of Field's account.

Finally, I describe a conflict that appears to Benacerraf between the requirements of a semantics for the language of mathematics and those of a plausible account of mathematical knowledge. Concentrating on the semantical side, I argue that no unified semantic theory should be expected. Nevertheless, an explanatory account of truth for mathematical statements is available, in the form of a theoretical definition of truth, as first formulated by Tarski. I maintain that the correctness of this theory of truth is not a genuine issue in the philosophy of mathematics. In particular, the theory does not entail or even lend support to Platonism.

Thesis Supervisor: George Boolos,
Professor of Philosophy

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Detailed acknowledgments have rightly been judged self-indulgently autobiographical, but equally they may provide the only human interest in an otherwise dreary treatise. When ones college and graduate education spans eighteen years, one inevitably incurs a multitude of debts. I am no exception. I owe my most general philosophical orientation to W.V.O. Quine, from whom I had the privilege of learning logic not once, but twice. Richard Henson introduced me to the philosophy of language and, perhaps unwittingly, first conveyed the idea that my philosophical potential might not be flatly ordinary. David Bennett introduced me to the philosophy of mathematics, and even more to the present point, suggested that my reluctance to accord truth and falsity to mathematical statements might result from thinking truth to be something it is not. He recommended Alfred Tarski, as therapy. Michael Friedman taught me without visible reward. Judith Thomson not only encouraged me to take up graduate studies, but found me a place as well. Patrick Winston and the M.I.T. Artificial Intelligence Laboratory sponsored my novitiate at this Institute. Hilary Putnam supplied the general structure of these investigations in his Locke Lectures and the seminar meetings that surrounded them. As important as what he said was what he left unsaid. And most importantly, Richard Cartwright and George Boolos provided the supervision, encouragement, patience, and guidance that enabled me finally to write a thesis. The one taught me how to treat of philosophy generally; the other, how to treat of philosophy of mathematics.

It was a particular disappointment to me that circumstances had taken Professor Cartwright out of the country before I had been able to bring this work to its culmination, for he had listened with persistence to a great many ill-formulated ideas, prodding and cajoling me until I got them straight. No one who has not studied with him can appreciate the range and power of his philosophical way.

For specific discussions in aid of the present project, I thank Rogers Albritton, Timothy Appelt, Richard Boyd, Joshua Cohen, S. Marc Cohen, Warren Goldfarb, Paul Horwich, Jeffrey Poland, Scott Soames, James Thomson, Judith Thomson, Barbara Von Eckardt, and belatedly, Joseph Bevando.

For assistance with typing and manuscript preparation, I thank Rowland Cannon, Jr., Diane Harris, John Magee, Maxine Pollard, and Joan Yatteau.

On a more personal level, my parents enabled me to continue this madness without incurring even more debts that I did. And Mary Ellen Sullivan and our girls have both loved me and bet their lives on me. My gratitude to them is most heartfelt of all.

And now I owe all of these people a public exculpation for blunders and falsehoods contained herein.

ONE:

SENSE ABOUT PROPER NAMES

In a thought provoking footnote to his important work, Naming and Necessity, Saul Kripke exempts from his criticism of the classical theory of proper names a proposal of Quine's that names correspond to a kind of predicate. Kripke claims that Quine's proposal leaves the project that he undertakes in his essay essentially unchanged. I want to investigate whether that claim is true. Quine's suggestion has not been as widely noticed as have other views on naming. Perhaps many share the suspicion Kripke enunciates that the proposal is not "a substantive theory of the reference of names."[1] Quine's own disclaimers reinforce that suspicion. One reason that I want to discuss Quine's proposal is that I think that it can develop enough philosophical flesh to provide a serious alternative. But perhaps a deeper reason is that I believe that considering the problems Kripke discusses with Quine's proposal in mind will undermine--to good effect--many of the analytical tools that Kripke employs in his discussion.

In the first part of this paper I will explain exactly what Quine has proposed. I will contrast a narrower with a broader construal of

[1] Saul A. Kripke, Naming and Necessity (Cambridge, Mass.: Harvard University Press, 1980), p. 29, n. 5; originally in Semantics of Natural Language, ed. Donald Davidson and Gilbert Harman (Dordrecht: Reidel, 1972), pp. 253-355.

the proposal and relate the contrast to issues about the nature of philosophical analysis. And I will criticize some of the detail of Quine's exposition. In the second part I will raise some of Kripke's questions about naming and reference on the supposition that names are predicative. After extending the notion of rigid designation in a natural way, I will try to generate doubts about certain widely accepted cases of rigid designation. Doing so will emphasize the importance that Kripke attaches to the question, "How is the reference of names determined?" In the third part I will criticize typical theoretical answers to this question and suggest that the question embodies dubious philosophical presuppositions. Finally I will suggest ways of reformulating the question that would enable Quine to answer it.

I. PROPER NAMES AND PREDICATES

Russell's theory of descriptions is a well-known proposal for explaining the logical form of sentences containing definite descriptions. Under the proposal such sentences are analyzed, to use the term that Russell himself used, into certain replacement sentences which exhibit their own logical structure. Russell claimed that the replacement sentence in each case expresses the same proposition as the original sentence, this being the standard of a correct analysis. The key feature of the proposal is that the replacement sentences contain no definite descriptions; rather, certain constructions out of the quantifiers of first-order logic and the identity relation do the work of the definite descriptions. The replacement sentence is not constructed by substituting an equivalent phrase for each definite description in the original. Russell denied that a correct analysis of sentences containing descriptions could be given this way. No definition of definite descriptions in isolation is possible; they are to be defined in context. Russell took himself to have satisfactorily elucidated the logical function of definite descriptions if he could give a systematic method of producing, for each given sentence, an equivalent sentence with no definite descriptions. He specifically denied that definite descriptions, as constituents of the sentences in which they occur, have meanings of their own. Instead, sentences containing them are seen to be meaningful and to have the meanings that they do have by considering the meanings of the replacement sentences when the analysis is completed. Stated thus in terms of meanings, these claims seem unreasonable. They are better understood rephrased as follows, keeping in mind Russell's doctrine of propositions. The proposition expressed by a sentence containing a description is not a function of the reference of the description. Such a sentence nonetheless does express a proposition, and we can tell what proposition it expresses by considering the proposition expressed by the replacement sentence issuing from the analysis.[2]

[2] The original appearance of the theory of descriptions was in Bertrand Russell, "On Denoting," _Mind_, 14 (1905), pp. 479-493; rpt. in his _Logic and Knowledge_, ed. Robert C. Marsh (London: Allen & Unwin, 1956), pp. 39-56. For a more narrowly focused presentation, see Bertrand Russell, "Descriptions," (ch. 16 of) _Introduction to Mathematical Philosophy_ (London: Allen & Unwin, 1919), pp. 167-180.

Without going into much of the detail of Russell's theory, let us notice how it applies to a simple example. The sentence,

(1) The teacher of Plato was bald,

is analyzed by the theory as,

(2) $(\exists x)(\forall y)((y \text{ taught Plato} \leftrightarrow y = x) \;\&\; x \text{ was bald})$.

The analysis shows that to assert (1) is to assert that something both taught Plato and was bald and, what's more, nothing else taught Plato. The replacement sentence (2), unlike (1), contains no definite descriptions; nevertheless, no part of (2) corresponds to the description, 'the teacher of Plato'.

In various of his writings, W.V. Quine has suggested that we use Russell's analysis as a model for eliminating from our discourse all singular terms, not only definite descriptions, but what philosophers call proper names as well.[3] The sentence,

(3) Socrates was bald,

is exactly like our earlier example save for having the proper name 'Socrates' in place of the description 'the teacher of Plato'. We

[3] Willard Van Orman Quine, "On What There Is," Review of Metaphysics, 2 (1948); rpt. in his From A Logical Point of View, 2nd ed., rev. (Cambridge, Mass.: Harvard University Press, 1980), pp. 1-19; see pp. 7-8 in particular. Methods of Logic, 3rd ed. (New York: Holt, 1972), pp. 230-234. Word and Object (Cambridge, Mass.: M.I.T. Press, 1960), pp. 176-190. Philosophy of Logic (Englewood Cliffs, N.J.: Prentice-Hall, 1970), pp. 25-26.

ordinarily suppose that each sentence attributes baldness to a particular person, as it happens, the same person in both cases. Whatever else the two expressions do in their respective sentences, in (1) 'the teacher of Plato' identifies that person, while in (3) 'Socrates' identifies him. Quine's suggestion is designed to preserve the similarity of logical form that (3) thus apparently bears to (1).

The Russell analysis of example (1) is an instance of the quantificational schema,

(4) $$(\exists x)(\forall y)((Fy \leftrightarrow y = x) \;\&\; Gx).$$

If he can find an analysis of (3) that also is an instance of that schema, Quine will have captured that apparent similarity; he will have shown that the two indeed share a single logical structure at some non-trivial level of analysis. To find the appropriate instantiation of the predicate letters of the schema, it is clear enough that we should begin by supplanting the predicate letter 'G' with the predicate 'was bald' just as before. But with what can we supplant the other predicate letter? The novelty of Quine's suggestion lies in his answer to this question. Supplant the schematic 'F', he says, with the predicate, '= Socrates', yielding,

(5) $$(\exists x)(\forall y)((y = \text{Socrates} \leftrightarrow y = x) \;\&\; x \text{ was bald}).$$

This maneuver is possible not only in this case, but in any other, because of the logical equivalence of any sentence '. . . a . . .' with '$(\exists x)(\forall y)((y = a \leftrightarrow y = x) \;\&\; \ldots x \ldots)$'

so long as 'a' has referential position in '. . . a . . .'. Since '$(\forall y)(y = a \leftrightarrow y = a)$' is a valid formula, from '. . . a . . .' we can derive '$(\forall y)(y = a \leftrightarrow y = a)$ & . . . a . . .', and from that, so long as 'y' is not free in '. . . a . . .', we derive '$(\forall y)((y = a \leftrightarrow y = a)$ & . . . a . . .)' by a well-known rule of passage. An existential generalization yields '$(\exists x)(\forall y)((y = a \leftrightarrow y = x)$ & . . . x . . .)'. On the other hand, if we have '$(\exists x)(\forall y)((y = a \leftrightarrow y = x)$ & . . . x . . .)', an existential instantiation yields '$(\forall y)((y = a \leftrightarrow y = b)$ & . . . b . . .)', and a universal instantiation yields '$(b = a \leftrightarrow b = b)$ & . . . b . . .'. The theory of identity gives us 'b = b' and thus 'b = a & . . . b . . .'. Since 'b' has referential position in '. . . b . . .', a substitution again licensed by identity theory and an easy truth-functional simplification give us '. . . a . . .'. We have proved the desired logical equivalence.

Quine bases his recommendation of this treatment of proper names on several technical advantages. One advantage accrues in what Quine calls singular inference. Proper names are commonly treated in systems of formal deduction as individual constants or, what comes to the same thing, in systems that do not have individual constants, as free variables. The usual rules of inference provide for instantiation of a universally quantified variable with an individual constant and for existential generalization of an individual constant. Of course these rules, like all deductive rules, are supposed to preserve truth. Unfortunately, the application of these rules can generate false conclusions from true premises when an individual constant is schematic for a proper name which happens not to name anything. From '$(\forall x)(x = x)$', an axiom of identity, one can derive 'Pegasus = Pegasus', and from that, '$(\exists x)(x =$ Pegasus$)$', which says that something is identical to Pegasus and which is evidently false. A somewhat more subtle deduction derives the false conclusion,

Sherlock Holmes lived on Baker Street,

from the evidently true premises,

$(\forall x)$(the detective adventures of x were described by Doyle $\rightarrow$ x lived on Baker Street),

and,

$(\forall x)$(x = Sherlock Holmes $\rightarrow$ the detective adventures of x were described by Doyle).

What is required is the instantiation of 'x' by 'Sherlock Holmes' in both premises as well as in the previously mentioned axiom of identity. These instantiations are licensed by the rule for individual constants given above. Then two applications of _modus ponens_ give the false conclusion. (Similar inferences with '0/0' as the problematic singular term are familiar mathematical fallacies.) In both examples what has gone wrong is that the proper name fails to be the name of anything, whereas the proof of soundness for quantificational proof procedures only guarantees them under all interpretations of the individual constants and other uninterpreted symbols. But an interpretation assigns something rather than nothing to every individual constant.

A careful treatment of proper names as individual constants requires that we assure ourselves that all names name something before applying the rules of inference for individual constants to them. Quine's proposal relieves the rules for individual constants of this care and instead makes necessary the explicit statement of a premise asserting the existence of something named by the proper name. In a Quinean rendering of the second example, 'Ba' does not represent the conclusion,

Sherlock Holmes lived on Baker Street,

and '$(\forall x)(x = a \rightarrow Dx)$' does not represent the premise,

If anybody was Sherlock Holmes, his detective
adventures were described by Doyle.

For both premise and conclusion more complicated schemata are required, such that the conclusion does not follow from the given premises without an additional premise, '$(\exists x)(\forall y)(y = \text{Sherlock Holmes} \leftrightarrow y = x)$'. Since the latter is false, the deduction countenanced by the usual rules of inference is no counterexample.

Quine's views on the philosophical point of the proposal as described so far enable him to avoid certain problems that arise when we put the proposal to work in service of broader philosophical purposes. Because of his impatience with the notions of meaning and synonymy and proposition, he has no interest in the question whether the replacement (5) means the same as (3) or whether it expresses the same proposition as (3). Something of that sort must hold for it to serve as an analysis in the traditional Russellian sense. Quine recasts the task of analyzing the puzzling or logically troublesome locutions in ordinary discourse into the task of regimenting, as he calls it, ordinary discourse into such a form that the logical puzzles do not arise.[4] This is but one example of his refashioning the philosophical program of Russell and the logical positivists to accommodate the quite devastating criticisms that appear in his early papers culminating in "Two Dogmas of Empiricism". Regimentation is constrained not by the preservation of meaning but by the fulfillment of the purposes that the ordinary locutions were originally advanced to serve, fulfilling them in language appropriate for scientific theorizing. Simplification of theory provides the rationale for regimentation in case after case, and economy in the rules of deduction, all the while keeping them rigidly mechanical, is the frequent criterion of simplicity. So goes his account of the virtue of the reformed treatment of singular inference that we have outlined, and so also goes his familiar rejection of constructions with referentially opaque positions. In a hopeful paragraph Quine leaves the question of whether the regimented locutions do serve the purposes

[4] See Section 33 of Quine, Word and Object, entitled, "Aims and Claims of Regimentation."

intended to the atheoretic judgment of the theorist. He writes,

> . . . on the one hand there is theoretical deduction and on the other hand there is the work of paraphrasing ordinary language into the theory. The latter job is the less tidy of the two, but still it will usually present little difficulty to one familiar with the canonical notation. For normally he himself is the one who has uttered, as part of some present job, the sentence of ordinary language concerned; and he can then judge outright whether his ends are served by the paraphrase.[5]

Purposes of scientific theorizing are the only purposes that Quine has much interest in serving. A consideration of the character of scientific theory as Quine sees it reveals why logical equivalence rather than any stricter equivalence usually guarantees adequacy of paraphrase. In his view most sentences in a scientific theory merely mediate between observation sentences. They can do so in a more or less elegant way, thus his present concern with simplicity, and they can do so with more or less violence to our inherited conceptual scheme, thus the concern with conservatism enunciated in other contexts. The job of mediating is done as well by a logical equivalent as by any given sentence, for the threads of mediation consist of logical relations: what other sentences imply and are implied by the given sentence. Since logical equivalence preserves these logical relations under substitution, replacing the ordinary sentence by a logically equivalent regimented sentence is guaranteed to yield a theory

[5] Quine, Word and Object, p. 159. See also the first paragraph lying wholly within p. 160.

as suitable as the given theory. The question of sameness of meaning, which Quine regards as spurious anyway, is completely avoided. In the case at hand Quine provides a proof of logical equivalence between the unregimented sentence containing proper names in any referential position and the regimented replacement in which proper names occur only to the right of the identity sign.

Quine's program of regimentation will not prove satisfying to any whose interests lie precisely in determining the logical structure of the unregimented sentences occurring in our ordinary discourse and in determining in particular the contribution of proper names to that structure. Supposing that our interest lies there, we want indeed to ask whether (5) is an analysis of (3). An argument for logical equivalence does not show that it is, for logical equivalence is too loose a relation to insure synonymy or, on most views, sameness of proposition expressed. All pairs of logically valid sentences (or of contradictions) are logically equivalent, yet only on extreme views do all valid sentences express the same proposition. One can show that the sentence,

If the moon is made of green cheese, then nothing is what it is,

is logically equivalent to the sentence,

The moon is not made of green cheese,

but no one who takes meaning seriously thinks that they mean the same thing. Perhaps (probably) when pressed we cannot effectively determine identity of propositions expressed or sameness of meaning. If it follows that we cannot give usable standards for what constitutes an analysis, we might more modestly ask whether the logical form of (3) is indeed that displayed by the schematic (4). Is the coarsest structure of (3) an existential quantification? And does the presence of proper names in referential position always mark a sentence as a particular kind of existential generalization?[6] All of these questions exceed the reach of anything that Quine argued for in his theory of names.

Taking these questions seriously immediately calls to our attention the fact that in (5) the proper name 'Socrates' is not eliminated at all; it still appears, this time after the identity sign. And in every case, carrying out the transformations indicated by the schematic proof of pages 12 and 13 merely moves the proper names to this characteristic position. We said that the point of Russell's theory of descriptions was to analyze descriptions away by producing equivalent sentences in which no descriptions occur. If we want an analysis of sentences with proper names that will explicate their logical function, we had better not be satisfied with a further batch of sentences themselves containing proper names.

Even though Quine shuns analysis, he addresses himself to the issue just raised because he seeks another benefit from his regimentation besides that of rationalizing singular inference. He employs

[6] In Essay Three (pp. 95-98, below) I raise doubts about even these seemingly more modest questions.

it in understanding existence claims. The logical structure of sentences affirming or denying existence has long puzzled philosophers. Making sense of denials has especially led to extravagances of philosophical theory.[7] The traditional puzzle is, how can the sentence,

(6) Pegasus does not exist,

state something that is true? If there is no Pegasus, then 'Pegasus' names nothing. So what is the sentence about? In no true case of a denial of existence is there anything for the relevant sentence to deny existence of. In <u>Word and Object</u>, Quine approaches the problem by way of asking whether 'Pegasus' has referential position in (6).[8] In false cases, analogues of (6) survive the test of substituting for the name other names or descriptions of the same entity. All alike are false. But if that position is referential, then 'does not exist' is a predicate and one with empty extension. We suppose that the truth of predications depends on (or at least goes with) membership of the subject of predication in the extension of the predicate. And then 'does not exist' cannot be true of anything and thus not of Pegasus.

Suppose we take (6) to be 'not (Pegasus exists)' and apply Quine's method of regimenting proper names to the fragment governed by 'not'.

[7] For Quine's amusing account of some of these, see "On What There Is."

[8] Pp. 176-177 and 179.

(Of course the proof of logical equivalence requires that the proper names have referential position, and we have just conceded some doubt about the present case, but let us ignore that for now.) This instantiation of the replacement schema produces,

$$\text{not } (\exists x)(\forall y)((y = \text{Pegasus} \leftrightarrow y = x) \;\&\; x \text{ exists}).$$

The second conjunct ('x exists') of the part of this sentence governed by the quantifiers seems simply redundant. If we drop it we get,

$$(7) \qquad \text{not } (\exists x)(\forall y)(y = \text{Pegasus} \leftrightarrow y = x),$$

which presumably is to be read as 'nothing is Pegasus', or 'There is no such thing as Pegasus'. This analysis of sentences like (6) does not originate with Quine, but he is happy to embrace it as a second benefit of his proposal. More needs to be said though, for the problem about referential position recurs with (7). 'Pegasus' has referential position in (7) as normally parsed, and so long as (7) is false that is not puzzling. But again, what if (7) is true? (I take it that it is true.) As Quine puts it,

> [T]here is just something wrong about admitting that 'Pegasus' can ever have purely referential position in truths or falsehoods; for the intuitive idea behind "purely referential position" was supposed to be that the term is used purely to specify its object, for the rest of the sentence to say something about.[9]

[9] Quine, Word and Object, pp. 176-177.

We have found two difficulties with the presence of proper names in the replacement sentences--an obstacle to the claims of Russellian analysis and a problem about referential position. Only the second of these worries Quine since he does not intend to explicate the role of proper names in ordinary discourse anyway, but it motivates a significant extension of his proposal.[10] Quine suggests that we not consider the predicate '= Socrates' (or '= Pegasus') to be composed of an identity sign followed by a proper name or to be composed of any logical materials at all. We are to take it as logically simple. The 'is' in 'is Socrates' is merely a copula. What appears as '= Socrates' in the regimentation is a monadic predicate which, for logical purposes, has no structure. The predicate happens to be true of Socrates and of nothing else. In some passages Quine suggests that we use an alternative formulation--"socratizes", "pegasizes"--to remind ourselves of this.[11] On this account there is no problem about the position of a singular term 'Pegasus' in the denial of existence (7), for no singular term occurs; '= Pegasus' is a predicate, i.e., a general term, and the

[10] Quine is not quite candid about this. Remember that singular inferences are guaranteed by the proposal even if proper names remain in the regimented sentences. A conclusion schema correctly representing a sentence whose proper names purport to refer can be derived only from premises with appropriate existential force. I must acknowledge a third benefit which Quine thinks quite important, the closing of truth-value gaps. This too requires the extension. I have chosen not to discuss it because the issues are similar to those discussed with respect to existence claims.

[11] Quine, "On What There Is," p. 8.

notion of referential position is not applicable.

It is time that our attention revert to the proof of the eliminability of proper names set out on pages 12 and 13. The careful reader will have noticed that the proof given is not exactly Quine's proof.[12] Oddly enough, Quine's own replacement schema is '$(\exists x)(x = a \& \ldots x \ldots)$'; the constructions that make for uniqueness have been omitted. I bothered to spell out the proof in order to show that they can be included. Our proof provides for a replacement algorithm exactly paralleling that of Russell for definite descriptions, and it is only slightly more complicated than Quine's proof. Our version has the virtue of preserving and explicating the similarity of logical structure between (1) and (3). So long as we care about attempting a real analysis, that seems a virtue worth embracing. Quine is interested enough in the ordinary meaning of proper names to acknowledge that his version does not carry over "the purport of uniqueness".[13] He justifies his version with a comparison to certain other general terms. 'Cousin of' expresses a relation that in fact is symmetrical, and we commonly appeal to its symmetry in the course of deduction. Whether the symmetry of cousinhood derives from the meaning of 'cousin' or from contingencies of nature is an open question not easily settled. It is unclear what considerations should count towards settling it. Quine regularly appeals

[12] Quine's own version is in Word and Object, p. 178.

[13] Quine, Word and Object, p. 182.

to economy of logical theory in these cases, and I am sympathetic to that appeal. But in the present case no economy is noted, and I am unable to discover one. Since that algorithm works as smoothly which generates a replacement sentence requiring uniqueness of reference of proper names for truth, it seems the better choice. Nothing Quine says undermines our intuition that the original (3) implies that there is exactly one Socrates as much as (1) implies that there is exactly one teacher of Plato.[14]

Concerning either proof--Quine's original or the revised version above--two objections need to be aired. Both objections arise as a result of Quine's parsing the replacement sentence so that '= Socrates' is a logically simple predicate. In the first place the proof appeals explicitly in its second half to the axioms of identity theory. The substitution of 'a' for a once existentially quantified free variable, yielding the schematic '. . . a . . . ', crucially depends on the identity of the referent of 'a' with the referent of that variable. Indeed it is that very substitution which requires the limitation to contexts in which the proper name has referential position. But if the predicate '= a' has no structure, then the symbol '=' occurring in it does not express the identity relation, and the axioms of identity are irrelevant. Taking it as the identity sign is simply an equivocation.

[14] Quine's exposition in Methods of Logic, p. 233, does require uniqueness. There the method of paraphrasing for proper names is conflated with the method for definite descriptions, using the iota operator, which implies uniqueness, for both.

And half of the proof no longer goes through when we avoid this equivocation.

The second objection is similar. Both halves of the proof treat 'a' in '= a' as an individual constant, which can be existentially generalized in the first half and which can occur in an instance of the substitution axiom of identity in the second half. Again, if '= a' is logically simple and contains no occurrence of a singular term, treating 'a' as an individual constant is an equivocation. This objection can be generalized to avoid depending on Quine's suggestion that '= a' be considered simple. Whether that predicate is simple or not, in the proof rules of inference for individual constants are applied to 'a', which is schematic for the proper name under consideration. The logical nature of proper names is just what the proof is intended to clarify. To assume that they obey the logical laws for individual constants is to beg the very question at hand.

It must be said in Quine's defense that it is far from clear what he intends the point of his proof to be. We have portrayed it as an argument for the logical equivalence of (3) and (5) and of all analogous pairs represented by the schemata of the proof. So understood, it is intended to certify the correctness of a proposed analysis of sentences containing proper names. The two objections quite conclusively deny it this function. Quine never explicitly assigns the argument this function. He says that the proof "shows that the purely referential occurrences of singular terms other than variables can be got down to the type '= a'," and later speaks of "a theorem of confinability of singular terms to the position '= a'."[15] In our most generous moods, we still must ask why such a theorem has any relevance in this context, failing the purpose of certifying the proposed theory of proper names.

It remains for us to consider the prospects of Quine's proposal as an analysis of the ordinary functioning of proper names in light of the criticisms of his argument that I have made. Only one strategy seems to be available. Since proper names are eliminated by the proposal only under the full version in which '= Socrates' is logically simple, we must use such unanalyzed predicates and try to explain what kind of predicates they are. And since the argument for logical equivalence is fallacious if '= Socrates' has no structure, we must get along without the argument. On the latter count we are left in as hopeful a position anyway as that from which Russell defended his analysis of definite descriptions. He offered no proof of the logical equivalence of (1) and (2). It is hardly surprising that no logical deduction can be found to prove the equivalence of an unanalyzed locution and a proposed analysis. To apply the apparatus of deductive logic to any sentence we must know beforehand what logical form the sentence has. Discovering that is precisely the task of analysis. The results of analysis can hardly be expected to guarantee the correctness of the analysis. Furthermore, we have argued that more is required for analysis than logical equivalence. More comprehensive arguments are needed to make Quine's theory a

[15] Quine, <u>Word and Object</u>, pp. 178 and 179.

convincing analysis.

How did Russell argue for the theory of descriptions? He offered various typically philosophical considerations. He did not base his theory on narrow linguistic or logical grounds; he employed it instead in a wide-ranging philosophical theory that had both epistemological and metaphysical aspects. It is true that the theory of descriptions has many specific logical advantages as well. Russell applied it very successfully to the problem of denials of existence involving definite descriptions. And it preserves the law of excluded middle for cases where nothing satisfies the definite description. The theory clearly handles the analogue of Frege's problem about the Morning Star and the Evening Star, where descriptions rather than names flank the identity sign. As a technical device the theory found application in Russell's logicist reduction of mathematics in Principia Mathematica. Though Russell evidently had hopes for the theory as part of an explanation of propositional attitudes and other cases of oblique reference--this is the kind of case that occupies most of the text of "On Denoting"--it is doubtful that an understanding of definite descriptions helps much here. So goes the scorecard on logical puzzles for Russell's theory. It is an impressive performance that has secured a permanent place in logical theory for Russell's analysis. But equally dear to Russell's heart, it accorded with and reinforced his emerging philosophy of logical atomism. It protected the atomist doctrine of propositional functions and propositions from apparent counterexamples. It made unnecessary a notion of being that, as Russell had come to believe, generated logical absurdities. It provided an explanation of our knowledge of entitites with which we have no primary epistemological contact.

Though Quine refuses to interpret his proposal as an analysis of proper names in ordinary language, a case can be advanced for it resembling that of Russell's in form. We have already seen several logical advantages, those that recommend it as a regimentation. A review of the puzzles about reference yields a score very close to that of the theory of descriptions. Quine's theory results in parallel solutions in every case, when proper names are substituted for descriptions. Even Frege's problem seems to be handled by the theory, though this case particularly awaits explanation, from the point of view of epistemology, of predicates like 'socratizes'.[16] More importantly, Quine's proposal too fits into a wide ranging philosophical theory. Quine clearly advocates a doctrine of the primacy of predication that accords with his epistemological and metaphysical views. His account of language learning and the development in each speaker of the apparatus of reference places predication and the learning of predicates in center stage. And he hopes for a theory of language committed only to the existence of predicates and other inscriptional entities. The unity of these views naturally depends on a predicative theory of names.

[16] Tyler Burge has offered arguments from the point of view of linguistics in support of Quine's proposal as a theory of ordinary proper names. See his "Reference and Proper Names," The Journal of Philosophy, 70 (1973), 425-439; rpt. in The Logic of Grammar, ed. Donald Davidson and Gilbert Harman (Encino, Ca.: Dickenson, 1975), pp. 200-209.

It is appropriate to undertake an investigation of this wider rationale for Quine's elimination of proper names. I will begin an investigation by contrasting this view with the most influential current alternative, the doctrine of Kripke and others that employs the notion of rigid designation and sharply distinguishes names, and perhaps natural kind terms, from other expressions. The distinction is usually drawn by attributing sense to the expressions in most linguistic categories but excepting the category of names. Quine would not be friendly to a dissent that insisted that names too have sense, but there are other ways than that to dissolve the distinction.

II. PREDICATES AND RIGID DESIGNATION

If we are tempted by a predicative theory of proper names, the question naturally arises whether proper names are rigid designators nonetheless. In "Identity and Necessity", Kripke suggested something very much like the following as "a simple intuitive test" for whether an arbitrary singular term is a rigid designator.[17] Where α is any singular term that denotes, consider the proposition expressed by $\ulcorner\alpha$ might not have been $\alpha.\urcorner$ If this proposition is true, then α is not rigid; otherwise α is rigid. (To avoid confusion, we must recognize

[17] Saul A. Kripke, "Identity and Necessity," in Indentity and Individuation, ed. Milton K. Munitz (New York: New York University Press, 1971), pp. 135-164; see pp. 148-149 in particular.

that, in the test sentence, the first occurrence of α must be taken to have wide scope. Understanding the sentence otherwise would trivially result in the judgment that no singular term is rigid. If the scope of the first α is narrow, the test sentence is equivalent to, $\ulcorner$possibly not $(\alpha = \alpha)\urcorner$, which expresses a falsehood so long as α denotes. I take it that the test sentence is roughly equivalent to the philosopher's sentence, $\ulcorner\alpha$ is such that it might have existed and yet not been $\alpha.\urcorner$) This test yields the expected result in the case, for example, of 'the Roman Catholic President of the United States'. The sentence,

> The Roman Catholic President of the United States might not have been the Roman Catholic President of the United States,

indeed expresses a truth, for Kennedy, i.e., the Roman Catholic President, might never have run for President at all. What is more, Al Smith might have defeated Hoover in 1928, and if as well Kennedy had never run for President, then Al Smith, not Kennedy, would have been the Roman Catholic President of the United States. Therefore, 'the Roman Catholic President of the United States' is <u>not</u> a rigid designator, just as Kripke would have expected.

Of course this result does not follow in the case of every definite description. On Kripke's view, certain other definite descriptions are shown by this test to be rigid designators. To use Kripke's example, the sentence,

> The positive square root of 25 might not have

been the positive square root of 25,

expresses a falsehood; it follows that 'the positive square root of 25' is a rigid designator.

A natural extension of this criterion enables us to apply the notion of rigid designation to general terms and even to arbitrary monadic predicates. (Here, perhaps, it would better be called, "rigid denotation".) The transition is eased by considering the general term produced by dropping the definite article of a definite description. Though the complex term, 'Roman Catholic President of the United States', is general, it happens to denote only one thing, that is, to have an extension with only one element. Since we have already noticed that this element, Kennedy, might not have been elected President at all, we know that someone who in fact was a Roman Catholic President might not have been a Roman Catholic President; in other words, we know that some Roman Catholic President might not have been a Roman Catholic President. We will say, therefore, that the term, 'Roman Catholic President', is not rigid. In contrast, it is false that some positive square root of 25 might not have been a positive square root of 25, because, as we have seen already, the only positive square root of 25--the number 5--must be a positive square root of 25. Accordingly, we will say that 'positive square root of 25' is a rigid general term. Nothing prevents us from using a similar criterion for general terms with wider denotation. Because some President might not have been President, 'President' is not a rigid term. And because it is false that some square root might not have been a square root, 'square root' is a rigid general term. Evidently, in the case of any general term β, if the sentence, ⌜Some β (thing) might not have been (a) β,⌝ expresses a truth, β is not a rigid term; otherwise β is rigid.[18, 19] Only the constraints of grammar prevent us from using this test in the case of any monadic predicate. Again we can use a philosopher's sentence and adopt the following test for rigidity of any monadic predicate, ϕ. If the sentence, ⌜Something such that ϕ might not have been such that ϕ,⌝ expresses a truth, ϕ is not rigid; otherwise ϕ is a rigid predicate.

Just as the intelligibility of this criterion requires that the notion of a thing's being β *essentially* makes sense, the notion of a rigid term has other close connections with essentialist doctrines. It should be obvious that for any term β that is rigid, the sentence, ⌜If a thing is β at all, then it is β essentially,⌝ expresses a truth. In a manner of speaking, if β is a rigid term, the property which β expresses is one which is essential to anything that has it at all. This connection is important because the ability successfully to recognize rigid

[18] We are committed, in setting out this criterion, to half of what Quine calls the third grade of modal involvement, for we are committed therein to the intelligibility of Aristotelian essentialism. We are not, however, committed to the legitimacy of quantified modal logic or even of unquantified modal logic, for nothing we do requires the iteration of statement operators.

[19] This criterion needs to be refined. In the case of a rigid term, we want to rule out the possibility of something which actually lies outside the extension of the term, counterfactually lying within it. A similar refinement is required for the notion of rigid designation itself.

designation would comprise a method of settling the question, "Just which properties are essential?", a method utilizing linguistic considerations. Perhaps it would not judge all cases; the connection with rigid designation does not constitute a necessary condition for a property's being essential, but it does constitute a sufficient condition. And it implies that in the case of a property which satisfies this sufficient condition, given that it is essential to something, it is essential to anything that has it at all. In the interest of theory, some might conclude that the condition is necessary as well. If so, all essential properties have this somewhat surprising character. If men are essentially rational, then anything that is rational is essentially so.

We have extended the notion of rigid designation to cover general terms and other predicates. What we now want to examine is the case of proper names, even supposing that they are predicative. Kripke claims that proper names are rigid designators. The argument that he offers several times and with some embellishment really reduces to this simple intuitive test. Kripke asks us, "Is it true that Socrates might not have been Socrates?" He claims that the obvious answer is, "No." He does not really argue for this, but simply claims that any of us who understand the proposition see right off that it is false.

In the interest of caution, perhaps we had better offer the strongest challenge we can to these intutitions. What kind of case can we imagine that will raise the most serious doubts about these intuitions?[20] I will describe the best one that I have been able to find after an effort of some years. It will be convenient to fictionalize the case somewhat, although I hope everyone can see that the fiction could have been avoided. Suppose that certain facts about Socrates are as I will describe. For all I know they are, although I grant that it is not very likely. Suppose that Socrates had an identical twin who did not live very long, who in fact lived only a few minutes or hours after conception. After the crucial division of the original zygote that resulted in the coming to be of both Socrates and his twin, the developing organism, Socrates, successfully implanted himself into the uterine wall, matured into a healthy live fetus, and eventually was born. On the other hand, the twin failed to implant himself and died within a few hours for want of sustenance. As it happened neither Socrates nor his mother nor anyone else even knew that Socrates had an identical twin because the twin died so shortly after conception. In order to understand the case I am drawing, it is crucial to remember that I am supposing these to have been the actual historical events. Though it is unlikely that these events did occur in the case of Socrates' conception, nothing that we know about Socrates rules it out. It is very likely that such events have occurred in the conceptions of

[20] There is an engaging case about a ship, derived from Hobbes, in Hugh S. Chandler, "Rigid Designation," The Journal of Philosophy, 72 (1975), pp. 363-369. Chandler's case poses a problem precisely because we are unsure about the identity of the ships. Identity is taken for granted in my case.

some human children, though of course no one will ever know which ones.

Why does this case pose a challenge to the intuition that it is false that Socrates might not have been Socrates? The reason is simply that Socrates' dead twin brother might have been Socrates, while Socrates himself might have died _in utero_. What I am imagining--the counterfactual situation that might have come about, given that the factual situation was as I have described it--is that the individual that was Socrates, rather than successfully implanting, might have died within minutes of conception. And the twin might have implanted himself, matured into a healthy fetus, been born in approximately 470 B.C., grown up in Athens, conducted philosophical discussions in the Agora, found an exceptionally bright student in Plato, been convicted of corrupting the youth of Athens and sentenced to drink hemlock, been described in the Platonic dialogues, and so on. In short the twin might have done everything that Socrates in fact did. I am supposing this counterfactual situation to be entirely indiscernible by human eyes from the actual historical situation.

Fair enough, you say. Things might have been just as I described them, but why do I suppose that the case would correctly be described as I originally described it, that is, as Socrates himself not being Socrates but his in fact dead twin being Socrates? My shortest reply would be, I suppose, that _that_ seems to me a correct way to describe this counterfactual situation. It is tempting to respond to intuitions with further intuitions. Intuitively, this seems to me to be a case in which not the actual Socrates, but another person entirely, would be

Socrates. But I can say more than that about the case.

The first things I want to say are intended to lay to rest certain misapprehensions you may have about what I am claiming. First, I am not claiming that the actual Socrates is identical in the counterfactual situation to his twin brother, who would have been Socrates. That is, to adopt the lately fashionable jargon, I am not claiming that the man who is Socrates in the actual world is identical to the man who would be Socrates--the actual Socrates' twin brother--in the possible world I have described. On the contrary, the actual Socrates is identical to the individual in the counterfactual situation who died within minutes after conception. And Socrates' twin is identical to the individual in the counterfactual situation who lived to be born and went on to teach Plato.

Second, I am not proposing some particular solution to a problem about transworld identity. I do not believe that there is a problem about transworld identity, and I agree with Kripke that descriptions of counterfactual situations take it for granted that there is no problem about the identity of an individual who is supposed, contrary to fact, to fit some description that he really does not fit. My very description of the imagined situation depended on this. I said, "Suppose that the individual who was Socrates had died young and the individual who in fact died young had lived and eventually drunk the hemlock."

Third, I am not denying that Socrates must be identical with himself. I claimed that Socrates might not have been Socrates and advanced a scenario to illustrate that claim. I did not claim and would deny that Socrates might not have been who he is. I have no idea what kind of scenario anyone might advance to convince us that such a thing is possible.

Fourth, and perhaps most obviously, I am not supposing that the sentence, 'possibly not (Socrates = Socrates)', expresses a truth. Just as I can reasonably suppose that the teacher of Plato might not have been the teacher of Plato without supposing that the sentence, 'possibly not (the teacher of Plato = the teacher of Plato)', expresses a truth, I can make the analogous supposition when the definite description is supplanted by a name. The scope distinctions that are commonly made for definite descriptions to resolve an apparent paradox can be carried over to the case involving names. Of course, this claim commits me to a proposition that Kripke denies, namely, that distinctions of scope apply to proper names as well as to descriptions.

It remains for me to explain why I think that my case is correctly described as one in which someone other than the actual Socrates would have been Socrates. It is worthwhile to notice that in this case the twin would have been named "Socrates". Everyone would have called him "Socrates," and the same people (with one exception!) as actually uttered and believed such sentences as the following would have uttered and believed them:

(1) Socrates is down in the market place wasting his time talking about virtue again.

(2) Did you hear that Socrates finally got what was coming to him?

(3) Socrates claimed that knowledge is virtue.

(4) What we know about Socrates largely depends on Plato's portrayal of him.

Of course I am not making the mistake of supposing that just this much settles the issue. As is often emphasized, Socrates might not have been named "Socrates" at all, even supposing that he actually was. And his best friend might have been named "Socrates". It is not necessary with regard to every counterfactual situation that we conclude that the person supposed to have the name "Socrates" be supposed to _be_ Socrates. To guard against this mistake it is sometimes asked whether, granting that people might have called the twin "Socrates", they would be using the word in the same way that we actually do. Would they be speaking the same language that we actually speak? I believe that this way of looking at the issue makes the answer I give nearly irresistible. I have supposed that things in my counterfactual situation were, with very slight exceptions, exactly as things actually have been. Everyone believes the same things. Everyone says the same things. Nothing distinguishes the two situations that could be known by anyone. The only difference between them is a slight variation in a microsocopic event occurring within the womb of Socrates' mother. Actually one of two very small organisms died while the other flourished; counterfactually, it was the first mentioned that flourished, while the other died. Supposing that this difference would make the language we speak a different language defies belief.

But, you might say, there is one crucial difference between the two languages; the person designated by the name 'Socrates' is different. Therefore, the name 'Socrates' is used differently in an absolutely crucial respect. Before accepting this objection, we will do well to ask why this respect is supposed to be crucial for names. No parallel claim would be made for general terms. Suppose that Socrates had never married and hence that the term 'husband' did not denote him; there will be no temptation to conclude that speakers in that case would be using the term differently, that they would be speaking a different language, even that one small corner of their language would be different. Someone might resist this analogy with the explanation that names, unlike the term 'husband', are rigid designators. In this context that explanation would beg the question at issue. Another explanation might be this. Only two characteristics of a singular term are relevant to its semantics--its reference and its sense. In the case at hand we have agreed that the reference of 'Socrates' in the language of the counterfactual situation is different from the reference of 'Socrates' in our actual language. If I insist nonetheless that the semantics of 'Socrates' are the same for the speakers in my counterfactual situation as for us, then I must attribute that sameness to a sameness of sense. I must disagree with Kripke and agree with philosophers like Dummett that proper names have sense.

III. REFERENCE WITHOUT SENSE

Before I discuss the issue of whether proper names have sense, let me raise it again from a different direction. I am ready now to retreat a bit. Actually I do not believe that if Socrates' twin had not died but had lived and done everything that Socrates in fact did, then he would be Socrates. I have no idea who in that case would be Socrates. And I have been trying to shake your confidence that you know who would be. I believe that there is no answering that question. There is nothing about the semantics of the name 'Socrates' in our language that settles in every bizarre imaginable situation who the name applies to. And, this being the case, Kripke's simple intuitive test fails to tell us whether 'Socrates' is a rigid designator or not. Our intuitions are not up to the task. I surmise that Kripke would respond that there must be an answer to the question, because if there were none, there would be no answer to another question that he asks in several places in Naming and Necessity, namely, "How is the reference of 'Socrates' determined?"[21] I surmise that Kripke believes that whatever mechanism determines what the reference of 'Socrates' actually is, also determines

[21] Kripke, Naming and Necessity, pp. 27-28, p. 29, n. 5, p. 32, and p. 106, among many others.

what the reference of 'Socrates' would be in every possible counterfactual situation.

I suspect that the question, "How is the reference of 'Socrates' determined?", gets its force from its similarity to the question, "How is the reference of 'the teacher of Plato' determined?" In the latter case the point of the question can be seen in the answer given. An account can be given, presumably, of how the various constituents--'the', 'teacher', 'of Plato'--each contribute to the description's picking out just the individual that it picks out. One way to remind ourselves that the reference of the description is a function of all these constituents together is to notice how the reference changes when we replace one or another of the constituents with a parallel expression. Whereas the description as it stands refers to Socrates, if we replace the word 'Plato' with the word 'Aristotle' (yielding 'the teacher of Aristotle'), the reference changes to Plato. If we replace the word 'teacher' with the word 'student' (yielding 'the student of Plato'), the reference changes to Aristotle. And so on. Perhaps we think that we can explain why it is that each of these descriptions has one rather than another thing as its reference. To say these things is simply to rehearse the Fregean doctrines that the sense of a composite expression is a function of the senses of its constituents, and that the reference of any expression is a function of its sense. From these propositions it immediately follows by composition of functions that the reference of a description is a function of the senses of its constituents.

The classical view of proper names--the Frege-Russell view--was one which tried to project some of the successes of this theory onto the case of proper names. Since the theory does its work in cases where compositionality obtains, it was tempting to suppose that proper names do have a hidden composition that is revealed by supplying the description which each is supposed to abbreviate. The supposition accorded well in Russell's case with an emerging theory of knowledge by description; the theory extracted Russell from a skeptical quandary over how he could know about any entity that did not bop his consciousness on its head. The initial implausibility of this theory of proper names was mitigated slightly by the Searle and Strawson versions, which loosened the connection of a proper name with this or that particular property. Then the compositional theory had to be much more elegant, and so much the better.

Once we abandon this attempt to put compositionality to work for the case of proper names, we no longer have this kind of answer to give sense to the question, "How is the reference of 'Socrates' determined?" Quines's proposal is that names are like descriptions in that they are fundamentally predicative. Kripke says that on this view the original question yields to the question, "How is the extension of 'Socratizes' determined?"[22] That is not a question that is commonly asked with regard to predicates generally. For some predicates, again those that are composite, it may be an account of their structure that is wanted.

[22] Kripke, Naming and Necessity, p. 29, n. 5.

The predicate, 'Roman Catholic President of the United States', has in its extension only Kennedy, and this fact can be inferred in part from information about the linguistic material out of which the predicate is built. But for syntactically simple terms--'blue', 'mountain', 'walks'--I do not know what kind of answer is wanted. According to the proposal before us, it is just this kind of predicate that corresponds to each proper name. 'Socratizes' and 'pegasizes' are syntactically simple and, for all I know, are semantically simple as well.

Of course some philosophers have claimed that many syntactically simple terms have semantic structure. This view appears in the semantic theory of Katz. Indeed it might seem plausible that 'brother' is semantically composed by conjunction out of 'male' and 'sibling'. But it seems as plausible that 'sibling' is semantically composed by disjunction out of 'brother' and 'sister'. What is simple depends on what primitives we begin with, and no generally accepted theory has selected a particular set of primitives.

In any case we are bound on anyone's theory to arrive at a set of simple terms, and it is not clear what is wanted in answer to the question, "How is its extension determined?", asked about one of these. Some readings of the question elicit responses of a different kind that seems expected, responses that leave semantics behind. "How do we know what the extension of such and such a term is?" Perhaps we do not. And even when we do, our ways of knowing are multifarious and ill-sorted. "Is there a function from term to extension?" Of course. It takes 'blue' to blue things, 'mountain' to mountains, 'walks' to things that are walking. I have roughly described the function in extension. If you want a definition of the function or a procedure for computing it, neither I nor anyone else can give you one.

But aren't I ignoring a classical answer to Kripke's question? The question, "How is the extension of 'mountain' determined?", may seem unproblematic because, according to the Fregean tradition, its extension is determined by its sense. Because Dummett adheres to this tradition wholeheartedly, he argues that even proper names have sense. And because Kripke notices that learning someone's name seems nothing like learning the meaning of a new word, he thinks that there is a special problem about how names get their reference. We need to ask of each why this invocation of sense is supposed to be illuminating. If there is a general problem about what determines the reference of a term, there is as pressing a problem about what determines the reference of the sense of a term.

The problem most apparently arises for what I call translational theories of sense. In these the sense of each expression in a natural language is an expression in some neutral language that is imagined to underlie the variety of natural languages--semantic markerese, or a language of thought. The sense of an ordinary term turns out to be another term, in a hidden language. What then determines the reference of these hidden terms?

In contrast there is another kind of theory of sense for which this problem is solved too easily. These are the possible worlds theories. In Russell's phrase, they get by theft what the others cannot get by honest toil. For them the sense of a term is a function from possible worlds to extensions. Plug in the actual world, or any other world you care about, and out pops the extension. About these we could ask not what determines the reference of the sense of the term, but what determines the sense of the term in the first place? Why is this particular function rather than another associated with the term at hand?

It is time to recall that we began this inquiry with an account of proper names suggested by Quine. He would deny, I suppose, that his simple predicates have sense. After all, he has repudiated the realm of meanings. Insofar as senses are supposed to be abstract entities such that each term of a language has its own particular sense, in virtue of which the term has the extension it has, Quine would deny that there are such things as senses. His view would be uniform with regard both to proper names and their associated predicates and to garden variety general terms. If Kripke's questions about how terms get their reference simply play advance man to this kind of answer, Quine would repudiate the questions. If instead what is wanted is an explanation of how speakers come to use language at all, how they associate particular terms with features of their environment, Quine has set forth the rudiments of an account in _The Roots of Reference_ and more recent papers. This account is fundamentally psychological. Whether or not Quine's particular psychological account is correct, if any psychological account turns out to satisfy the kind of questions that Kripke constantly raises in _Naming and Necessity_, it will undoubtedly fail to tell us, in every imaginable counterfactual situation, whether this or that, is or is not, denoted by a term. Accordingly it will provide no foundation for the notion of rigid designation. That is the moral of my story about Socrates and his twin. And I suggest that these considerations, perhaps among others, motivate Quine's skepticism about essential properties.

TWO:

FIELD'S THEORY OF TRUTH

In a widely admired recent paper entitled, "Tarski's Theory of Truth," Hartry Field sets out two versions of a definition of truth of the sort first proposed by Alfred Tarski in the early nineteen-thirties.[1] In it Field provides one version as an account of Tarski's actual definition. On the other hand he recommends his own version as the definition (or characterization, as he prefers to call it) Tarski should have given in order to make transparent exactly what philosophical work Tarski's theory does. He claims that his version is less "misleading" and less subject to a particular "misinterpretation" (p. 348). The feature of the usual philosophical elaboration of Tarski's definition that he regards as a misinterpretation is the claim that the given definition should be acceptable to someone suspicious of semantics because it defines truth entirely in other than semantic terms. It is Field's opinion that the Tarski theory, properly understood, reduced the notion of truth to other semantic notions. If Field's opinion proved well-founded, it would certainly be an important

[1] Hartry Field, "Tarski's Theory of Truth," <u>The Journal of Philosophy</u>, 69 (1972), 347-375. All further references to this work appear in the text. Tarski's original article is Alfred Tarski, "The Concept of Truth in Formalized Languages," in his <u>Logic, Semantics, Metamathematics</u>, trans. J. H. Woodger (Oxford: The Clarendon Press, 1956), pp. 152-278; partial rpt. in <u>The Logic of Grammar</u>, ed. Donald Davidson and Gilbert Harman (Encino, Cal.: Dickenson, 1975), pp. 25-49.

realization because Tarski himself doubtless thought it a primary virtue of his theory that it defined truth without using other semantic notions. In virtually every one of his many papers on the subject, Tarski cited this feature of his theory; he perhaps thought it the feature most responsible for the theory's philosophical significance. I will argue that Field is mistaken in his interpretation of Tarski, that Tarski had good reason to avoid the version that Field prefers, that Field's version cannot have the kind of generality that he wishes for it, and that Field's hopes for completing the version he recommends by going on to give acceptable definitions of the more basic semantic terms are unfounded.

I. TWO VERSIONS OF TARSKI'S DEFINITION

Field calls his own version of the truth definition, "T1", and the version which he attributes to Tarski, "T2". Because we will have to attend to certain details of the definitions, there is no avoiding setting them out as Field did. I do this in a tabular format in order to make evident where the versions agree and where they differ. The clauses left blank in T2 go over from T1 without change (pp. 350 and 354).

	T1	T2
(A) 1.	'x_k' denotes$_s$ s_k.	1.
2.	'c_k' denotes$_s$ what it *denotes*.	2. 'c_k' denotes$_s$ $\bar{c}_k$.
3.	⌜$f_k(e)$⌝ denotes$_s$ an object a iff	3.
(i)	there is an object b that e denotes$_s$, and	(i)
(ii)	'f_k' *is fulfilled by* <a, b>.	(ii) a is $\bar{f}_k(b)$.
(B) 1.	⌜$p_k(e)$⌝ is true$_s$ iff	1.
(i)	there is an object a that e denotes$_s$, and	(i)
(ii)	'p_k' *applies to* a.	(ii) $\bar{p}_k(a)$.
2.	⌜~e⌝ is true$_s$ iff e is not true$_s$.	2.
3.	⌜e_1 & e_2⌝ is true$_s$ iff e_1 and e_2 are true$_s$.	3.
4.	⌜$(\forall x_k)(e)$⌝ is true$_s$ iff for each sequence s* that differs from s at the kth place at most, e is true$_{s*}$.	4.
(C)	A sentence is true iff it is true$_s$ for some (or all) s.	(C)

Readers familiar with the notation and theory of predicate logic will not find these definitions as opaque as will others. A few remarks about the definitions will suffice not to supply those others with the necessary background, but to enable one to decide which class he falls in. The definitions are not general. They define truth for a particular language. They divide the sentences of that language into two sorts, those that are true and those that are false. The language under consideration has variables, individual constants, function symbols, and monadic predicates. The expressions of each of these kinds are ordered by a correspondence with the positive integers; the numerical subscript comprising a part of each such expression reveals, in each case, what the corresponding integer is. The language includes certain other symbols--'~', '&', and '∀'-- usually called logical constants, and the symbols '(' and ')', which are used to indicate grouping and scope. The language includes no expressions except those that can be built out of the fundamental expressions just surveyed by repeated applications of familiar syntactical operations. The truth definitions, in the more or less perspicuous form in which they are set out, are inductive definitions that mirror the possible syntactical operations so as to yield a truth value for every grammatical sentence of the language. However, the application of a standard device of set theory would convert them to another (even) less perspicuous form, in which they would have the virtue of constituting normal (Tarski's term), or explicit (Field's term), or proper (Putnam's term) definitions. The idea here is that a proper definition of a (monadic) predicate is of the form depicted by the schema '$(\forall x)(\phi\ x \leftrightarrow \psi\ x)$'. In that schema '$\phi\ x$' is to be replaced by the predicate being defined. And '$\psi\ x$' is to be replaced by an open sentence of some language in which the definition is being given; the open sentence must have only one free variable, namely 'x'. Inductive definitions, as is apparent in the present case, do not have this form.

The definitions employ a metalinguistic variable 's' which ranges over sequences, the items of which (s_1, s_2, etc.) are whatever objects the language under consideration talks about. These sequences must be at least as long as the number of variables in the language (Tarski actually calls for infinite sequences), but that imposes no restriction on the number of objects over which the sequences range, for nothing prevents an object's occurring more than once in one sequence, nor again, requires that each sequence exhaust all of the objects there are.

The only terms of the definitions remaining to be remarked upon are those which are peculiar respectively to T1 and T2. I have underlined three terms in T1--'denotes', 'is fulfilled by', and 'applies to'; they comprise the more basic semantic notions in terms of which T1 defines truth. Of course these terms do not appear in T2 since it purports to define truth entirely in other than semantic terms. T2 does include expressions consisting of words of the language for which truth is being defined overscored by a horizontal bar. A full discussion of these expressions would be rather lengthy. It is enough for our purposes to say that they are "abbreviations for the English expressions that are translations of the corresponding words of the language under consideration" (p. 354).

A casual reader of T2 might wonder whether it accurately represents Tarski's definition of truth. It is well known that Tarski's insight consisted largely in defining truth by way of defining satisfaction, yet T2 does not mention satisfaction at all. Tarski recognized that sentential connectives and quantifiers operate not only on sentences but on expressions with free variables as well. Only when he developed the notion of satisfaction could he account for the contribution of these open constituents of a sentence to its truth value. However, as Field mentions in an aside, the notion of $true_s$, which T1 and T2 in their different ways define inductively, captures Tarski's notion of satisfaction. Satisfaction is a relation which a sequence does or does not bear to a possibly open sentence. A sentence, whether open or closed, is $true_s$ just in case the sequence s satisfies the sentence in Tarski's sense. In both T1 and T2 truth is simply the limiting case of satisfaction for an expression with no free variables, that is, for a (closed) sentence. As it happens, if any sequence satisfies a sentence, all sequences do. This results from the peculiarities of clause (B)4--the same in T1 and T2--which specifies the effect of placing a variable-binding universal quantifier before an expression open in that variable. According to that clause, satisfaction of an expression by a sequence survives the binding of one of its free variables only if the expression is satisfied not only by the given sequence, but by any sequence produced by varying the relevant member of the given sequence over any object in the domain. Thus when all variables in an expression have been bound by quantifiers, yielding a sentence, it no longer matters what sequence was under consideration initially. That sequence satisfies the closed expression, that is, the sentence, just in case any other sequence does. Tarski's account of satisfaction is more explicit in certain respects than Field's schematic T2. Crucial issues will turn on just how satisfaction is defined for atomic formulae, so Tarski's account merits careful attention. But our current project demands that we provide ourselves with a broader view of Field's various claims.

We begin a comparison of T1 with T2 by noting what they have in common. As we said earlier, both T1 (as it stands) and T2 define truth only for a particular language. Neither provides an account of truth-in-a-language that ranges over all languages, or even more than one language. Field acknowledges this. He also acknowledges that the language for which truth is defined has a particular property which is not shared by most languages. To use Tarski's words, as quoted by Field, in this language "the sense of every expression is unambiguously determined by its form" (p. 348).[2] This unfortunate locution is intended to exclude two quite common linguistic phenomena. Excluded are words which are ambiguous properly speaking. There are no expressions in the language which, as we would ordinarily say, have more than one meaning. An example Field uses is the common noun 'bank', which means on one hand a certain kind of financial institution and on

[2] Tarski, "The Concept of Truth," p. 166

another, a sloping margin of land along a river. (There are several other hands as well.) Also excluded--this is the second and separate phenomenon--are expressions which are used more or less systematically to refer to different entities on different occasions of use. Here Field's examples are the pronoun 'I' and the given name 'John'. In these and in other typical cases, it is inaccurate to use the term 'ambiguous' because no variation in meaning is involved. In fact, as ordinarily understood, it is a feature of the meaning of 'I' that it can be used by each different speaker to refer to a different entity, namely the speaker herself. Nevertheless, each example does contribute to the possibility that on different occasions of utterance of a single sentence containing it, different statements, perhaps with different truth values, are thereby made. Presumably Field wants to exclude as well words like 'now' and 'there' and 'tomorrow' and 'immediately' for which it is at least doubtful whether the notion of reference is appropriate, but which similarly contribute to the dependence of the statement made by a containing sentence on time, place, speaker, audience, and other circumstances of utterance or inscription.

Because we will need to return to these matters, it will be helpful to have some terminology. Following another author, I will call <u>complete</u>, sentences of the sort included in the language for which truth is defined by T1 and T2.[3] Those sentences excluded by the consider-

[3] Richard Cartwright, "Propositions," in Analytical Philosophy, ed. R. J. Butler (Oxford: Basil Blackwell, 1962), pp. 81-103; see p. 97.

ations of the previous paragraph I will call incomplete.[4] Roughly my criterion is this. A sentence is complete just in case only one statement can be made by assertively uttering the sentence. If it is correct, as I believe it is, that the primary bearers of truth and falsity are statements, it comes as no surprise that truth with respect to sentences can be defined only for complete sentences. In these cases we can take the truth value of a sentence to be that of the single statement that can be made by uttering the sentence assertively. Since many different statements with different truth values could be made by uttering an incomplete sentence, any choice of one or another of these statements to transfer its truth value to the sentence would be arbitrary.

Though T1 and T2 have in common the features just discussed, they do differ in important respects. Field believes that the differences are such as to give T1 three advantages over T2. (1) Although T1 as it stands defines truth only for a language all of whose sentences are complete, Field claims that it can be extended without difficulty to cover incomplete sentences as well. He even outlines the extension; in it truth is defined not for sentences, but for tokens of sentences. In contrast, he says, T2 cannot plausibly be extended to cover incomplete

[4] It is perhaps unfortunate that Field uses these terms to distinguish different kinds of sentence token (p. 351). Since his distinction is not invoked herein, no confusion should result. We do need terms to mark our distinction. Complete sentences are roughly what Quine calls eternal sentences, but the connotations of that term are inappropriately narrow.

sentences. (2) Although T1 and T2 as they stand both apply only to a language whose vocabulary is fixed, Field claims that T1 can be reformulated beforehand to apply to the language as it acquires indefinitely many new items of vocabulary. Again he outlines the reformulation, and again he denies that a comparable reformulation of T2 is possible. (3) The third advantage is best stated as a disjunction. T1 clearly and unashamedly defines truth in terms of more basic semantic notions. Field seems to claim that either T2 does so as well but pretends otherwise, giving T1 the advantage of perspicuity, or else T2 does define truth entirely via nonsemantic notions, as advertised, but does so in a scientifically unacceptable way.

Eventually we will consider each of these claims individually. First we will do well to set out an important fact about T1 and T2 which anyone might suspect, but which Field does not acknowledge until very late in his paper. The fact is that, given suitable bridge definitions, T1 and T2 are virtually equivalent, though not quite logically equivalent. If we call these bridge definitions B (we shall see presently exactly what they are), we can state the logical situation as follows. $\ulcorner$B & T1$\urcorner$ is logically equivalent to $\ulcorner$B & T2$\urcorner$. Though T1 and T2 are not strictly logically equivalent, if we accept B we must believe that T1 and T2 are alike in truth value. And if B is logically true, then T1 and T2 are logically equivalent. Since B turns out to be pretty trivial, one might wonder what all the fuss between T1 and T2 is about. It is worthwhile to keep this question in mind during the discussion to come.

Indeed, when the logical relations between T1 and T2 are put this way, a good deal of Field's rhetoric about them is seen to be quite hollow. Field says in a number of places that T2 minus T1 is the trivial B (pp. 363, 364, and 370). He offers several analogies to explain this remark but ultimately bases it on the correct observation that the conjunction of T2 with B implies T1. Thus, he claims, T1 is a weaker version of Tarski's theory. But the situation is quite symmetrical, for the conjunction of T1 with B implies T2. Logical arithmetic is less like real arithmetic than Field would have us notice. The point of these remarks of Field's is to convince us that whatever interest there is in T2 there is also in T1. He needs this conclusion to refute the presumption that T2 must be more interesting and powerful than T1 since it does not appeal to undefined semantic terms. The presumption is false, Field argues, because what we have to add to T1 to make it imply T2, namely B, is devoid of interest. Setting the question of interest aside, we should remember that if B is true, even though it is trivial, then $\ulcorner T1 \rightarrow T2 \urcorner$ is true also, and thus T2 correctly defines truth if T1 does. They capture the same property. And the more like a truth of logic B turns out to be, the closer T1 comes to actually implying T2, so that the identity of the extension of 'true' in T1 and T2 is no accident. In order to deny the truth of B, Field will have to give up his argument that T1 is merely a more cautious theory than T2 and one our scientific scruples should prefer. As we will see, Field really advocates a modified form of T1 which embodies very strong semantic assumptions and which does not follow from

Tarski's theory at all.

Whether Field believes that B is true or that B is false, then, turns out to be a crucial question for evaluating Field's claims. So similar are T1 and T2 that the clever reader will already have guessed what B is, even if she has read neither Tarski nor Field. Since only three clauses of the alternative definitions differ, and these are the ones in which T1 employs its three semantic terms, all that is needed to make T1 and T2 equivalent to one another is to provide appropriate definitions of the three semantic terms. And since the number of simple individual constants, function symbols, and monadic predicates is in each case finite, we can provide such definitions. In the case of 'denotes', the following definition carries (A)2 of T1 over into the corresponding clause of T2.

(D) $(\forall e)(\forall a)$(e is an individual constant that <u>denotes</u> a $\leftrightarrow$ $((e = \text{'}c_1\text{'} \; \& \; a = \bar{c}_1) \vee (e = \text{'}c_2\text{'} \; \& \; a = \bar{c}_2) \vee \ldots \vee (e = \text{'}c_n\text{'} \; \& \; a = \bar{c}_n)))$[5]

I will not bother to give them, but analogous definitions (F), of 'is fulfilled by', and (A), of 'applies to', carry clauses (A)3(ii) and (B)1(ii) of T1 over into the corresponding clauses of T2. Then B is simply the conjunction of these definitions, ⌜D & F & A⌝. Tarski actually envisioned such definitions of various semantic relations, as Field himself observes. In his main paper Tarski specifically mentioned

[5] Compare Field's (D2) (p. 370); also, his (DE) and (DG) (p. 365).

definitions of the relation of denotation and of the relation of a monadic predicate defining a property. His definition of denotation was slightly different from (D) because it used what for him was the fundamental notion--satisfaction. But (D) is exactly what Tarski's concept of denotation comes down to for the case of individual constants.[6] It is because (D), (F), and (A) are specifically fashioned to transform T1 into T2 and vice versa that I have called them bridge definitions.

Now, one might ask, aren't bridge definitions unexceptionable? Can't we stipulate that 'denotes', 'is fulfilled by', and 'applies to', used in a theoretical setting, are defined any way we like, and thereby insure the truth of the definitions? How can Field deny them? It seems indisputable that we can, and adopting this line leads us to conclude that given B, i.e., given that we stipulate that B is true, T1 and T2 are merely different forms of the same definition. Neither is weaker nor stronger than the other, and only superficial convenience can make the form of one preferable to the form of the other. Even Field's argument that "the standard of extensional equivalence doesn't guarantee an acceptable reduction," and thus does not guarantee the equal acceptability of T1 and T2 has no force here (p. 363). On this

[6] Tarski, "The Concept of Truth," pp. 193-4 and p. 194, n. 1. It is unclear just what Tarski means by "name" in the footnote, where the definition of denotation is given, since there are no individual constants in the formalized language he considers, but only free variables. Unlike (D), the definition he gives would work as well for definite descriptions or other singular terms as for individual constants.

line B is very much like a truth of logic; one inevitably recalls Carnap's meaning postulates. If B is a meaning postulate, then T1 and T2 do not merely have the same truth value, but are logically equivalent.

There is a reply available to Field. He must grant that we can give these as definitions of three technical terms that only serve to recast T2, the definition we begin with, into a definition that uses the same words as T1. But then he must insist that this is not what he means by T1. When he uses these three semantic terms, he does not adopt them as terms of art to be given what meaning we like; rather he uses them to mean what they mean independent of any stipulation. When he says "denotes," so his reply goes, he is talking about denotation, and whether (D) captures the relation of denotation cannot be settled by stipulation.

When we consider (D) in that light, it becomes perfectly clear, despite his saying that it is trivially true, that Field does not believe that (D) is true at all.[7] He must deny (D) in order to extend T1 to secure the first advantage which he claims for it. In the extension which he sketches, Field attributes truth to certain tokens of incomplete sentences. He claims that tokens of words like 'I' and 'John' denote whatever the persons producing these tokens refer to by so

[7] Field actually says not that (D) is trivially true, just that it is trivial and "of no interest" (pp. 363 and 370). But he does not say that it is false.

doing. Thus he believes that the relation of denotation includes in its course of values such pairs as a particular token of 'I' together with Field himself or a particular token of 'John' together with John Adams. This is incompatible with (D) because the "only if" half of (D) limits the course of values of denotation to those pairs actually listed in (D), and the pairs listed have only word types as first members. No word tokens occur in the domain of the denotation relation at all according to (D), yet Field's extension presupposes that at least some word tokens denote. That is why Field must deny (D) in order to save the proposed extension. Similarly Field's further reformulation of T1 to secure his second claimed advantage runs afoul of (D). This reformulation enables T1 automatically to handle accretions of new vocabulary to the language by supposing, among other things, that new individual constants denote certain objects. Again the reformulation succeeds only if the denotation includes in its course of values pairs not listed in (D), and thus again, Field must deny (D) to save the reformulation.

It is interesting to notice the logical situation that obtains when (D) is weakened to accomodate these reformulations of T1. When we change the connective 'if and only if' in (D) to a mere 'if', we get,

(D') $(\forall e)(\forall a)(((e = \text{'}c_1\text{'} \ \& \ a = \bar{c}_1) \vee (e = \text{'}c_2\text{'} \ \& \ a = \bar{c}_2) \vee \ldots \vee (e = \text{'}c_n\text{'} \ \& \ a = \bar{c}_n)) \rightarrow$ e is an individual constant that denotes a).

(D') does not contradict Field's two proposed extension of T1. Given only

$\ulcorner D' \,\&\, F \,\&\, A \urcorner$, T1 and T2 are no longer equivalent. However, $\ulcorner T1 \rightarrow T2 \urcorner$ does follow from $\ulcorner D' \,\&\, F \,\&\, A \urcorner$. This suggests that T1 is actually the stronger theory, rather than T2, as Field claims. Intuitively speaking, T2 treats of a specific, though infinite, stock of sentences and effectively divides them into truths and falsehoods. On the other hand T1 employs the notion of denotation, along with the other semantic relations, to attribute truth and falsity to an open-ended supply of sentences. Though we are supposed to understand the relation of denotation, Field provides no effective method for determining when the relation obtains. In committing itself to such a relation, T1 constitutes the stronger theory.

It might be objected that we have misconstrued Field's argument for the modesty of his claims. Whether Field believes B or not is not at issue, so the objection goes, because it was Tarski who committed himself to something like B in suggesting definitions of the various semantic relations. What Field says is that T1 is a weaker version of Tarski's semantic theory, and Tarski's theory is not merely T2, but $\ulcorner B \,\&\, T2 \urcorner$. T1 does follow logically from that conjunction. As far as I can see, this point is of no interest whatever. What matters is whether Field's semantic theory is a weaker version of Tarski's theory. We have seen that $\ulcorner \sim B \urcorner$, as well as T1, must be part of Field's theory, and $\ulcorner B \,\&\, T2 \urcorner$ certainly does not imply $\ulcorner \sim B \,\&\, T1 \urcorner$.

II. TRUTH FOR SENTENCE TOKENS

What are we to say about Field's extension of T1 to cover incomplete sentences? No doubt he is right to say that no plausible extension of T2 will enable it to handle these cases. If his proposal for extending T1 were plausible, then it would constitute a genuine advantage. I now want to argue that it is not plausible. Field's account of the extension is not very full; for that reason my objections depend on making certain assumptions about the details of it. In order to avoid the obvious obstacles to attributing truth and falsity to incomplete sentences (i.e., to incomplete sentence types), Field changes T1 from a definition of truth for sentences to a definition of truth for tokens of sentences. Apparently the change applies to all sentences, not just to incomplete ones; under the revised T1, truth is a property of tokens of complete sentences as well as tokens of incomplete sentences. As we have noted, the definition assumes a relation of denotation between tokens of names and appropriate objects. Field says, "a name token denotes an object if the person who spoke or wrote the token referred to the object by so doing" (p. 351). Given that the revised definition attributes truth not to complete sentences but to their tokens, this latter claim seems to apply not only to names (so-called) like 'I' and 'John' which occur in incomplete sentences, but to names in complete sentences as well.

My first objection arises from this assumption. If we take Field at his word, the denotation of tokens of names turns out to be what recent literature has called speaker's reference, that object referred to by a speaker in using a referring expression.[8] This notion of speaker's reference was developed to handle certain pathological cases of reference. Due to some kind of misinformation a speaker occasionally refers to an object by using a description that does not correctly describe it or a name that is not a name for it. Donnellan's original example had a speaker at a cocktail party using the phrase, 'the man drinking a martini', to refer to someone who in fact had water in his glass. Kripke suggested that a similar phenomenon could occur with people's names. Smith is seen at a distance and taken to be Jones. In asking, "What is Jones doing?", the speaker refers to Smith. If these cases are correctly described in this way, there seems to be no reason why a similar case could not occur with an individual constant. I suppose that the least doubtful cases of individual constants in English are symbols like '1' and 'π', which name mathematical entities, but it would be awkward, though possible, to contrive an example using such a term. Since the characteristic of individual constants that is crucial for these matters is uniqueness of reference, let us suppose that 'Santa Claus' is an individual constant. (I doubt that it is a name that is given to many people.) It suits this context because it is easy to imagine a situation where a speaker who is misinformed uses the name to refer to someone. Suppose that in the course of describing his Christmas season visit to the department store, a child says, "Santa Claus had a soft white beard." Presumably he did refer to

[8] Keith Donnellan, "Reference and Definite Descriptions," Philosophical Review, 75 (1966), 281-305. Also see remarks in Saul A. Kripke, Naming and Necessity (Cambridge, Mass.: Harvard University Press, 1980), p. 25 and n. 3; originally in Semantics of Natural Language, ed. Donald Davidson and Gilbert Harman (Dordrecht: Reidel, 1972), pp. 253-355.

someone in using the name 'Santa Claus'; he referred to the man dressed up as Santa Claus, on whose lap he had sat. According to the revised T1 designed to handle tokens of incomplete sentences, the token of the sentence, 'Santa Claus had a soft white beard', that he uttered, was true, for that man did have a soft white beard. Unfortunately the sentence itself (the type) is not true. Since the sentence is complete, theories like T2 and the original T1 can handle it, and both would assign it falsity. Other theories would assign it no truth value. But no theory would assign the sentence itself truth. If the notion of speaker's reference is coherent at all, it should be obvious that examples can be developed in which a sentence itself is false on all accounts though the token is true on Field's theory. So it turns out that Field's extension of T1 assigns truth values to tokens of incomplete sentences at the expense of denying complete sentences that definiteness of truth value that was to be their chief virtue. Of course Field could avoid this result by leaving T1 intact for complete sentences and supplementing it for incomplete sentences in the way he suggests. But this would be to utilize two different relations of denotation in the definition of truth.

Further problems attend an attempt to assign truth values to tokens of sentences so long as truth is made to depend on what the speaker refers to in using a name token. According to Field's extension the following sentence token is false because there is no object that the person who wrote it referred to by using the name 'Santa Claus'.

Santa Claus had a soft white beard.

That result might not trouble us until we realize that the token would be false even if there were a Santa Claus. The person who wrote it merely displayed it as an example; he did not even try to use the name to refer to anything. Even more troubling is the result that the following token turns out to be true, not because there is no Santa Claus, but because the part governed by '~', being part of an example in which no reference was intended, is not true.

~Santa Claus had a soft white beard.

Field's extension must at least be augmented with a distinction between tokens used assertively and others. Obviously a theory of truth for sentence tokens must provide for truth-value gaps since tokens are used for all kinds of purposes beside asserting truths. Adequate machinery for making this distinction will so far exceed anything provided in either of T1 and T2 that is unclear that T1 has a significant head start.

It might be supposed that all of these objections arise merely from taking Field too literally about what the speaker refers to in using a name. Just as the original T1 considered only the semantic reference of names in complete sentences, perhaps the extension of T1 should shun the complexities of speaker's reference and employ the semantic reference of names in incomplete sentences, taking advantage of the variation of semantic reference from token to token. Though the suggestion sounds promising, there is reason to doubt that there is any semantic reference for many of the substantives that typically occur in incomplete sentences. The semantics of pronouns impose some constraints on their reference, but even in context the constraints are not strong enough to fix the reference. For many occurrences of 'he', 'she', 'it', and even 'you', an understanding of the speaker's subject matter and what the speaker is likely to want to refer to is required for resolving pronominal reference. And clearly nothing in the semantics of demonstratives like 'that' or 'those' will be adequate for determining the reference. For such denoting expressions speaker's reference is very likely all that there is on which to build a theory.

Speaking more broadly, Field's proposal for incomplete sentences errs in trying to attach truth values to sentence tokens at all. I observed earlier that truth attaches primarily to statements. A semantic theory of truth is possible for complete sentences only because of a certain simplicity of fit between those sentences and what statements can be made with them. The relations between incomplete sentences and the various statements made with them on various occasions are more complicated. The day might come when we understand these relations and have both a theory of truth for statements and an account of what statements are made on the occasion of uttering what sentence tokens and of what occasions result in no statement being made at all. Here there is plenty of room for Austin's observation that utterances can suffer many infelicities besides falsehood. When that day comes, there will be no obstacle to assigning truth values to certain sentence tokens parasitic on the truth values of statements made by uttering them. To do so now would be to mask ignorance with confusion.

III. GENERALITY IN TRUTH DEFINITIONS

In attributing the second advantage to T1, Field claims for it a kind of generality not enjoyed by T2. The form of T1 inevitably ties it to a language which is static, fixed once for all in the expressions it includes and the sentences that can be built out of them. It is a familiar feature of natural languages that they change over time. Specifically their vocabularies become ever larger as speakers encounter new things they wish to speak of and conceive new ideas they wish to express. Field suggests reformulating T1 to make it a theory of truth not only for the language under consideration at one moment, with the vocabulary and expressive power it has at that moment, but a theory that would continue to distinguish the true sentences from the false as the language grew in vocabulary. The reformulation achieves this by relieving the various clauses of T1 of their specificity. For example both in the original T1 and in T2, clause (A)2 goes schematic for a long conjunction, each conjunct of which mentions a particular individual constant and gives its denotation$_s$ for the sequence at hand. In the case of T2, saying what the various constants denote$_s$ requires different words in each case, so the conjunctive form of the clause cannot be avoided. In order to formulate a definite conjunction, the theory must settle on a finite stock of individual constants, one to be mentioned in each conjunct. By contrast each conjunct in T1 has the same predicate: such and such a constant denotes$_s$ <u>what it denotes</u>. This obviously evokes a generalization: any individual constant denotes$_s$ what it denotes.[9] The generalization accomodates any individual constant which accrues to the language in the course of time. Similar generalizations reformulate clauses (A)3 and (B)1;

[9] Compare Field's clause 2 (p. 353).

in each case specific mention of the various function symbols and predicates is discarded in favor of generalizations for any function symbol and for any monadic predicate. Apparently Field envisions a theory that will even accommodate novel expressions for negation, conjunction, and universal quantification (p. 353, n.10).[10] Since additional expressions for these logical operations do not enhance the expressive power of the language, this accommodation is not very important.

Before commenting on the reformulation, I want briefly to explore the question of generality in truth definitions. No doubt all philosophers would be gratified to have a theory of truth that was completely general. In principle--that is, in abstraction from the facts--a general theory could define truth _simpliciter_. In doing so it would yield a principle that no matter what s is, s is true just in case $\phi(s)$, where '$\phi(s)$' stands for the result of predicating of s, some monadic predicate that could be as complicated as the theory would require. A very simple consideration defeats the hope for such a theory. Nothing prohibits a single sentence from occurring in more than one language, and furthermore, from being true in one language and false in another. This can happen even with a complete sentence. The intuitive idea is that the meaning of the sentence differs in the different languages so as to make it true in the one and false in the other. Since the notion of truth _simpliciter_ is consequently doomed, any correct theory must employ the notion of truth-in-

[10] Field concludes the note with this remark, "The use of semantic categories in the generalized truth characterization raises important problems which I have had to suppress for lack of space in this paper." Perhaps he there anticipates some of the objections that I raise later in this section.

a-language that is good for all languages. This kind of theory would yield a principle that for any language L, no matter what s is, s is true in L just in case $\phi(s,L)$, where this time '$\phi(s,L)$' stands for the result of predicating of the pair, $\langle s,L \rangle$, some diadic predicate, again as complicated as the theory would require. Now Tarski knew very well that hopes for a general theory even of this form are illusory. Many languages --in fact all of the important natural languages--have a characteristic, which he called universality, which stands in the way of consistently applying the notion of truth to their sentences. In universal languages sentences can be formulated which speak of their own truth and falsity, and thus in these languages the Paradox of the Liar can be formulated.[11]

Our hopes must settle on an even more modest generality in a theory of truth. Nothing in Tarski's discussion argues against the possibility of developing a theory of truth which yields a principle that for all languages L, if L excludes semantic terms which apply to its own expressions, then no matter what s is, s is true in L just in case $\phi(s,L)$. Nonetheless, though he never explicity argued against this possibility, the theory of truth that Tarski developed did not have this form. We have constantly noted that Tarski's theory merely defined truth for a

[11] As is now widely understood, mere formulation of the Paradox of the Liar is not enough to establish a contradiction. Only in a theory that includes extralogical axioms can any proposition that is not a truth of logic be established. And only in a theory which has axioms or definitions which govern the notion of truth and which have a certain minimal power can the admission of the Liar sentence lead to contradiction. There are ways to avoid paradox even though the Liar sentence be considered well-formed, but they involve a considerable complication of the treatment of truth. See Saul Kripke, "Outline of a Theory of Truth," The Journal of Philosophy, 72 (1975), 690-716.

particular language. An examination of T2 reveals the extent to which the definition which it expresses is limited to the particular language under consideration. The three clauses peculiar to T2 each employ expressions--those overscored by the horizontal bar--which are translations of terms of the language for which truth is being defined. Nothing is said about how these translations are achieved or about what standards of correctness govern them. It is simply assumed as a prerequisite to the theory that correct translations of the relevant terms of the language can be expressed and that in constructing the definition of truth for the language we have these translations ready at hand. In itself the assumption is harmless enough because when we set out to define truth for the particular language we probably do have these translations at hand; if we did not, we very likely would not be interested in a truth definition for the language anyway. In the case of the very simple language for which Tarski actually gave a truth definition in his principal paper, the assumption is particularly harmless because that language is a part of the language we speak. Nevertheless, the important point with regard to generality in truth definitions is that it cannot be assumed that translations of all expressions in all languages are available. No method of generating translations for languages generally is even contemplated by Tarski. The apparatus which T2 uses serves only for defining truth-in-that-language for which it does define truth.

On the other hand, T1 does not employ translations of the language under consideration. In it there are no terms overscored by a horizontal bar. We saw Field exploit this fact in generalizing T1 to accommodate extensions of vocabulary. With a little tinkering the main part of T1 can constitute a diadic predicate to replace '$\phi(s,L)$', as required for a general theory of truth. Among other revisions, the basic semantic notions and the intermediate inductive notions used in T1 must all be relativized to L. Thus our old friend clause (A)2 becomes,

(A) 2. Any individual constant in the vocabulary of L denotes$_s$ in L what it denotes in L.

The question can hardly be suppressed whether T1 thus ultimately revised might serve as the hoped-for general definition of truth. Field carefully avoids any consideration of this question though I see nothing else that could motivate his demand for a scientifically acceptable definition of primitive denotation. Let there be no uncertainty about my answer to this question. T1 so revised cannot serve as a general theory of truth.

Some of its deficiencies are quite superficial and can therefore be ameliorated. I have in mind the limitation to monadic predicates, the treatment of only two of at least sixteen possible truth functions, and the treatment of universal quantification to the exclusion of the existential and other more unusual quantifiers (for example, 'exactly one', 'at most fifty-nine', etc.). We know enough about the logical relations among these different devices to expect that a sufficiently general account can be developed to handle them all. Actually some interesting problems do arise even at this superficial level. For example since there is no limit in principle to the number of places which a predicate in some language might have, no finite number of clauses like (B)1 will suffice. Some general clause must be constructed to handle n-place predicates for any number n. The clause must incorporate a general description of the sentence that results from supplanting the n different openings in an n-place predicate by the singular terms e_1, e_2, ... , and e_n respectively. It is common for an account of the semantics of a formal language to utilize such a description for that particular formal language, whose syntax is known. But it is unclear how to construct a description that will handle predicates in many different languages with different syntactic structures.

Tl suffers other deficiencies, these so profound as to make me skeptical that a general theory of truth can be hoped for at all. An outline of them has the present purpose of illuminating Field's more modest claim to generality for Tl, namely, the accommodation of an expanding vocabulary. My first worry concerns the definiteness of the proposed theory. Generalizing Tl to a variable language requires, as we saw, relativizing the basic semantic notions. The definition of truth now depends on a general notion of denotation-in-L good for all languages. Obviously the same considerations about generality previously rehearsed for the case of truth apply here to denotation. Is denotation-in-L to be different for each value of L, perhaps defined by a list of its course of values for each L? Of course this will not do, for the same reason that T2 would not serve as a general account of truth; it would require having beforehand translations of all of the names in all languages. What is required is a general account that says that e denotes a in L just in case $\psi(e,a,L)$. Of course this calls to mind Field's demand for a scientifically acceptable definition of denotation. But our current position is much worse than Field's position with the original, modest Tl. It is not just that we have no reduction of a fairly well understood relation; we have no idea for variable languages what the relata of denotation are.

Think of all the things that philosophers have claimed to be mediated by the relation of reference. Sentences, with truth values. Singular terms in oblique contexts, with their customary senses. Parts of paintings, with objects in the scene being depicted. Truth-functional connectives, with truth functions. Notes in a musical score with pitches. The physical states of complexes of neurons, with--I choose the phrase deliberately--anything you can think of. Our intuitive idea of the relation between given names and their bearers collapses under this burden.

I have other worries about the adequacy of anything like T1 as a general account of truth. These arise from doubts about the universality of predication as a vehicle for expressing truths. We are all familiar with opaque contexts of singular terms. In these the term seems not to denote at all, and thus the surrounding expression does not constitute a predicate. Some cases that seem to require unacceptable ontological commitments may best be viewed as not involving ordinary predications at all. 'The hole in this doughnut is small' may be true without there being doughnut holes, but not so long as it implies that something is small. Since these worries would carry us far afield, let us set them aside in favor of observing, now more clearly, Field's second advantage for T1.

If we do not understand denotation for variable languages, do we understand any better that notion as it applies to arbitrary extensions of the vocabulary of our particular language? I am not sure that we do. In order for Field's own reformulation of T1 to have definite content, we must assure ourselves that we understand the phrase, "what it denotes," for any individual constant that accrues to the vocabulary to the language. Field's theory has some protection when extensions of the vocabulary go so far as to shake our confidence, for his reformulation of T1 requires "that the general structure of the language be fixed, e.g., that the semantic categories (name, one-place predicate, etc.) be held constant" (p. 353). Perhaps the semantic category comprising individual constants, what Field calls "names", is characterized by at least a rough criterion of denotation for its elements. My worries about predication can similarly be forestalled with a characterization of the category of predicate, though I fear this will involve a circularity. If fixing the general structure of the language were to dissolve these worries, I would press no specific objection to the reformulation and would grant Field the advantage he claims for T1. I would only emphasize how modest an advantage it is. Any extension of the vocabulary of the particular language which will survive what has turned out to be a very strong qualification on the scope of the reformulation is such that we can see right off how to rewrite T2 to accommodate it.

I have said very little about Field's third advantage for T1, the one that has been most widely discussed by others. It should be obvious that I believe that Tarski was serious in claiming that his theory defined truth entirely in other than semantic terms. Tarski did not base his theory on general semantic relations, of which we have an adequate pretheoretic understanding though no theoretical account, precisely because he did not think there were any such relations. That leaves him open to Field's accusation that Tarski's truth is a scientifically unacceptable notion. Though I am unsympathetic to the physicalist doctrine with which he supports this accusation, I do think there is something to his worries. In order to carefully assess his accusation we must distinguish senses in which a notion can fail to be a scientific one. Tarski's truth is scientifically acceptable in that it is well-defined, or at least as much so as the terms in the language to which it applies. Precise definition is not enough, however, to give a term a place in certain kinds of scientific explanations. Certain explanations might best be called causal explanations; these are particularly prominent among scientific explanations. My own approach has been to suggest that truth, and reference, are not casual explanatory notions. If I am correct, there is an important sense in which they are not scientific notions and cannot be at the heart of scientific explanations.

THREE:

TRUTH AND THE PHILOSOPHY OF MATHEMATICS

An extended but severely critical scrutiny of another philosopher's writing usually answers to one of two motivations. Most often the critic profoundly respects the views being examined and wishes to explain why in the end he thinks them incorrect, even if tempting. The second motivation appears more rarely. Here the critic has little or even no respect for the object of his attention, but believes it necessary for the improvement of philosophical understanding to expose errors that are both egregious and generally unrecognized. Examples are painful to recall. Because my own motivation fits neither of these descriptions, I have hesitated to cast the remarks below in the form of an extended critical commentary--in this case, a commentary on Paul Benacerraf's article, "Mathematical Truth."[1] My discussion may seem somewhat diffuse, but Benacerraf's essay itself, proposing as it does a program of investigation for all of the philosophy of mathematics, is unavoidably diffuse and varied in content. There are conceptual difficulties facing Benacerraf that vitiate the thesis that he propounds, general and elusive as Benacerraf deliberately makes that thesis. I have not resisted structuring my own essay around Benacerraf's because his essay has structured my thinking on this subject

[1] Paul Benacerraf, "Mathematical Truth," The Journal of Philosophy, 70 (1973), 661-679. All further references to this work appear in the text. That paper was presented at a symposium on the subject of mathematical truth held at a joint meeting of the Eastern Division of the American Philosophical Association and the Association for Symbolic Logic, in December, 1973. The other symposiasts were Saul Kripke and Oswaldo Chateaubriand. Benacerraf read an earlier (1967) and longer version of the paper to audiences at several universities. That version circulated in typescript, and I have benefited from reading it in addition to the published version.

during the time since I first read it. If I am excused at all, it is by the likelihood that I am not unusual in this respect. I surmise--these things are surprisingly hard to verify--that the article has been widely read and discussed, and that the dilemma that it poses has begotten more than one recent treatise in the philosophy of mathematics.[2] The difficulties that confront Benacerraf have confronted others working in the philosophy of mathematics and in the philosophy of language. Perhaps our time will be well spent in clearing some of them up.

In the course of this study, I defend three primary claims regarding the philosophy of language and its relation to the philosophy of mathematics:

(1) No unified and comprehensive semantic theory for the language of mathematics is to be expected on the basis of current achievements.

(2) Nevertheless, a systematic theory of truth for mathematical statements is available, in the form of a theoretical definition of truth of the sort first formulated by Alfred Tarski, together with supporting axioms and rules of inference.

(3) The correctness of this theory of truth is not a genuine issue in the philosophy of mathematics. In particular, the theory does not entail or otherwise lend support to Platonism.

[2] Mark Steiner's Mathematical Knowledge (Ithaca: Cornell University Press, 1975) seizes one horn of the dilemma that Benacerraf poses, which will be described below. And Hartry Field's recent book, Science without Numbers (Princeton: Princeton University Press, 1980), seizes the other.

My defense of these claims emerges from a discussion of Benacerraf's analysis of the condition of contemporary philosophy of mathematics.

Benacerraf's analysis emphasizes a conflict which he claims to find between the requirements of a satisfactory semantics for the language of mathematics and the requirements of a plausible account of mathematical knowledge. I describe this conflict in Section I. In setting out his semantical requirement, Benacerraf unsystematically intersperses five notions: the notion of a theory of truth, the notion of a theory of logical form, the notion of an account of the semantics of singular terms, the notion of an explanation of the semantics of the quantifiers, and the notion of a formulation of the truth conditions of a proposition. Contrary to the presupposition of much work in semantics, I believe that these notions are more or less independent of one another, and that some of them are perhaps not even philosophically significant. I argue for this and thereby defend my first claim, that we should expect no comprehensive semantic theory, over the course of several sections. In Section II, I argue that no definitive theory of logical form is in sight, and that in fact there is no such thing as <u>the</u> logical form of a statement. In Section III, I maintain that theories of the semantics of singular terms cannot contribute significantly to the theory of truth. I argue in Section IV that a theory of truth does not explain the semantics of the quantifiers and that no other theory does either. An understanding of quantifiers depends on the general understanding of a language in which quantifiers occur and a grasp of the principles of reasoning which govern them. And in Section V, I point out that the call for truth conditions, unless it is to be answered trivially, presupposes metaphysical or linguistic doctrines which are not widely held. In each of the sections I take care to make clear what notion I am discussing and to distinguish it from other semantic notions with which it is sometimes confused.

Having distinguished it from other semantic notions, I turn in Section VI to the theory of truth for mathematics and the defense of my second claim, that we do have a systematic theory of truth for mathematical statements. I distinguish the notion of a theoretical definition of truth from the notion of an arbitrary statement of truth conditions. Then I urge that Tarksi's theory of truth qualifies as a theoretical definition of truth that elucidates the concept of truth for philosophical purposes. What's more, the theory has consequences that express what we ordinarily regard as most central to the notion of truth, namely, that a sentence is true if and only if the situation that it describes actually obtains; these consequences attest almost irresistibly to the correctness of the theory. I believe that these considerations should lay to rest whatever doubts Benacerraf has about the availability of a philosophically adequate theory of mathematical truth.

Running through Benacerraf's paper is the suggestion that embracing a Tarski theory of truth for mathematical language commits us to Platonism in the philosophy of mathematics. Thus the relevance of my third claim, that the correctness of Tarski's theory of truth is not a genuine issue in the philosophy of mathematics. Of course, what I mean is that, if the theory is properly understood, there should be no resistance to accepting it from philosophers with whatever views about the nature of mathematical objects and the proper forms of mathematical reasoning. The simple reason is that this theory of truth is neutral with regard to these metaphysical issues, just as Tarski intended it to be. The detailed defense of this third claim occupies Section VII. I formulate several principles that are central to Platonism. A careful examination of each of them shows it to be entirely independent of an at most mildly reformed version of Tarski's definition of truth.

This leads to what is perhaps the most interesting conclusion of this study, that contrary to widely held opinion and despite our habits of expression, there is no problem about the nature of mathematical truth. The important classic positions in the foundations of mathematics were not positions on the nature of mathematical truth, and even today the nature of truth is not at the heart of disagreements in the philosophy of mathematics. There are genuine disagreements, and questions involving truth enter into some of them, but there is no room for significant disagreement about what it is that we are saying about mathematical propositions when we say that they are true.

I. BENACERRAF'S DILEMMA

There appears to Benacerraf a fundamental dilemma in the philosophy of mathematics, a tension between the requirements of the semantics of mathematics (better, of mathematical language) and those of the epistemology of mathematics. True to his title, he takes his dilemma to obstruct a satisfactory account of mathematical truth. He divides philosophies of mathematics into two categories, what he variously calls "Platonistic" or "standard" views and what he calls "combinatorial" views, claiming that Platonistic (or standard) views are motivated by a concern for a satisfactory semantics and that combinatorial views are motivated by a concern for plausible conclusions in the epistemology of mathematics. The weakness of each kind of view coincides with the strength of the other. Thus if mathematical statements share their semantics with statements from the rest of our language (as on the standard view), we cannot account for our knowledge of them. On the other hand, if our knowledge of mathematical statements is accounted for in the natural way (as on the combinatorial view), their semantics must be divergent. Either way we lack an acceptable comprehensive view of mathematics, which must have both semantic and epistemological components. Therein lies the dilemma.

Whatever we discover below about the concern for semantics, the gist of Benacerraf's epistemological claim is that Platonism defends a realm of mathematical objects of which human beings could never know. According to a causal theory of knowledge, knowledge of any object is gained by entering into causal relations with it. Mathematical objects are presumed to be abstract and outside the causal nexus; for that reason nothing about them would be knowable if they did exist. Since we could know nothing about them, there is, according to this theory, no reason to suppose that they do exist. In contrast, combinatorial views in the philosophy of mathematics emphasize the role of proofs in yielding our knowledge of mathematical propositions, presumably meriting the blessing of causal theories of knowledge. However pressing these epistemological concerns are, I will not comment further on them. Our interest will lie in Benacerraf's concern for semantics.

Among the views that he calls combinatorial, Benacerraf includes Hilbert's formalism; the (unnamed) view that "the truth conditions for arithmetic sentences are given as their formal derivability from specified sets of axioms" (p. 665); the claim that "the Peano axioms are . . . 'analytic' of the concept of number" (p. 665); conventionalist accounts; and his own view, advanced in another well-known paper, that numerals are not names.[3] This does not seem to leave much to populate his first category; indeed he says, "Many accounts of mathematical truth fall under this [the combinatorial] rubric. Perhaps almost all." (p. 676) The one specific example of a Platonistic view that he gives is that of Gödel.[4] In calling the views in his first category "standard", perhaps Benacerraf means to suggest that it is this philosophy of mathematics that most of us, who are not in the grip of one or another odd theory, subscribe to or, perhaps, that most mathematicians subscribe to.

As is well-known, three positions dominated the classical controversy over the foundations of mathematics during the first third of this century. Of them, Benacerraf calls Hilbert's view combinatorial, but he mentions neither intuitionism nor logicism. We will consider some doctrines included in intuitionism in good time. For now it seems to fit comfortably in neither category. Though it clearly does not suffer epistemological failings of the sort Benacerraf finds in so-called standard views, I cannot see how it can sensibly be called a combinatorial view. Logicism, on the other hand, can fall into either category, depending on how it is embellished. Benacerraf scorns a kind of logicism in "What Numbers Could

[3] Paul Benacerraf, "What Numbers Could Not Be," Philosophical Review, 74 (1965), 47-73.

[4] This view is defended in Kurt Gödel, "What Is Cantor's Continuum Problem," revised version in Philosophy of Mathematics, ed. Paul Benacerraf and Hilary Putnam (Englewood Cliffs, N.J.: Prentice-Hall, 1964), pp. 258-273.

Not Be," and that seems to throw it into the Platonist category. This is the kind of (really a descendant of) logicism that portrays numbers as particular sets and then goes on to propound a Platonistic philosophy of set theory. In contrast, the logical positivists saw logicism as successfully reducing arithmetic to logic and then gave a conventionalist account of logical truth. Their brand of logicism thus seems to be combinatorial, on Benacerraf's telling. If this ambivalence seems puzzling, remember that logicism, that is, the classic view of Frege in the _Grundgesetze_ and of Russell in _Principia Mathematica_ and in his articles of the period of _Principia Mathematica_, claims simply that mathematics is a part of, or can be reduced to, logic. To the extent that Frege and Russell advanced views on the nature of logical truth in these writings, their views had very little in common.

Benacerraf's difficulty in specifying members of his first category of philosophies of mathematics, the so-called standard or Platonistic views, reflects an even deeper running difficulty in clarifying the notion of a satisfactory semantics for mathematics, a concern for which supposedly motivates these views, to the exclusion of right-minded epistemological concerns. What is the first horn of Benacerraf's dilemma? And what is Benacerraf's conception of a semantics for mathematics? He initially mentions "the concern for having a homogeneous semantical theory in which semantics for the propositions of mathematics parallel the semantics for the rest of the language" (p. 661). This characterization leaves unstated what Benacerraf takes to be included in a semantic theory. We might try to infer what is required by looking at a formulation of the semantics for the rest of the language, but Benacerraf tells us that he is "indulging here in the fiction that we _have_ semantics for 'the rest of language,'" or at least in the fiction that proponents of the standard view think that they have "such semantics, at least for philosophically important segments of the language" (p. 661, n. 1). Evidentally a homogeneous semantic theory is not something readily available.

A few pages later Benacerraf gives his full account of this horn of the dilemma as a "condition" on a comprehensive philosophy of mathematics. He describes it as "the requirement that there be an over-all theory of truth in terms of which it can be certified that the account of mathematical truth is indeed an account of mathematical _truth_" (p. 666). A few sentences later Benacerraf reformulates this condition, as follows.

> Another way of putting this first requirement is to demand that any theory of mathematical truth be in conformity with a general theory of truth--a theory of truth theories, if you like--which certifies that the property of sentences that the account calls "truth" is indeed truth. (p. 666)

These passages clearly indicate that Benacerraf's semantical condition on an adequate philosophy of mathematics requires an account of mathematical truth, that is, an account of truth in respect of the sentences, or propositions, or whatever bears truth and falsity, of arithmetic and the rest of mathematics. Equally clearly though, Benacerraf requires something more of an adequate philosophy of mathematics than just this. What is not clear is what it is that he requires in addition. Somehow the account of mathematical truth must accord with more general philosophical claims about truth, in order to certify mathematical truth as genuine truth. Thus begins what seems to me the fairest and most intelligible interpretation of what Benacerraf first requires of the philosophy of mathematics; in Section VI we will return to this theme and elaborate it considerably. For now I will merely reprove describing this desideratum as a semantics for mathematics that parallels the semantics for the rest of language. Philosophers beginning with Tarski have had reason enough for calling the theory of truth (a part of) semantics, but we will see how misleading this practice has been.

Benacerraf's additional remarks make it unclear that his first requirement is as I have just now construed it. Let us survey some of his other characterizations of his semantical requirement. Shortly following our last quotation, in evident even if hesitant elaboration, Benacerraf says,

> Perhaps the applicability of this requirement to the present case amounts only to a plea that the semantical apparatus of mathematics be seen as part and parcel of that of the natural language in which it is done, and thus that whatever semantical account we are inclined to give of names or, more generally, of singular terms, predicates, and quantifiers in the mother tongue include those parts of the mother tongue which we classify as mathematese. (p. 666)

Benacerraf has quite unconsciously switched from talking about a theory of truth for mathematics to talking about accounts of the semantics of singular terms, predicates, and quantifiers in our language generally and in our language as it is used to formulate mathematical propositions. Moreover he has apparently switched in the belief that no change of subject has occurred. He seems to believe that a theory of mathematical truth is the same thing as a semantical account of singular terms, predicates, and quantifiers. It remains ahead for us to consider what accounts of the semantics of these various categories of expression consist in. It will become clear that I believe that such accounts form no part of a theory of truth, and that I am inclined to believe that a theory of truth forms no part of any of them.

In the process of illustrating possible responses to his semantical requirement, Benacerraf makes other remarks that equally raise questions about what would constitute a satisfactory semantics for mathematics and about the relations between various semantic notions. To focus discussion, Benacerraf displays three sentences, one whose subject is geography, one whose subject is arithmetic, and one that is schematic:

(1) There are at least three large cities older than New York.

(2) There are at least three perfect numbers greater than 17.

(3) There are at least three <u>F</u> <u>G</u>'s that bear <u>R</u> to <u>a</u>.

Benacerraf first asks, "Do [(1) and (2)] have the same logicogrammatical form? More specifically, are they both of the form (3)?" A few lines later he asks, "What are the truth conditions of (1) and (2)?" Apparently in reply to this, he soon says, "[I]t seems clear that (3) accurately reflects the form of (1)." But then, evidently intending a contrast, he asks, "But what of (2)? May we use (3) in the same way as a matrix in spelling out the conditions of <u>its</u> truth?" (All pp. 663-664.) Benacerraf seems to be suggesting either that the notion of the logical form of (1) and of (2) is the same thing as the notion of the truth conditions of (1) and of (2) or else that some common answer can be given to questions about the logical form of (1) and (2) and about the truth conditions of (1) and (2).

Having asked whether we may use (3) "as a matrix in spelling out the conditions" of the truth of (2), Benacerraf goes on to illustrate replies to this question. He first says, "Some . . . have shied away from supposing that numerals are names and thus, by implication, that (2) is of the form (3)." Next, a few sentences later, he says of Hilbert, "[H]e did not regard all quantified statements semantically on a par with one another. A semantics for arithmetic as he viewed it . . . would certainly not treat the quantifier in (2) in the same way as the quantifier in (1)." And finally, still in illustration of possible replies, Benacerraf says, "On other such accounts, the truth conditions for arithmetic sentences are given as their formal derivability from specified sets of axioms." (All pp. 664-665.) These passages, coming on the very heels of one another, seem to conflate several notions. Dropping the theory of mathematical truth altogether, the passages quite unsystematically intersperse talk of the logical form and the truth conditions of sentences and of the semantics of singular terms and of quantifiers.

Many other passages from the article can be cited to demonstrate that Benacerraf's articulation of a semantical requirement on the philosophy of mathematics combines the notion of an acceptable account of mathematical truth with at least four other distinct notions, summarized as follows:

(1) The notion of a theory of the logical form of sentences whose subject is arithmetic, specifying in particular the logical forms of those sentences that say that there are numbers satisfying certain conditions.

(2) The notion of an account of the semantics of numerals and other singular terms in the language of arithmetic.

(3) The notion of an explanation of the semantics of the quantifiers in the language of arithmetic.

(4) The notion of a formulation of truth conditions for an arithmetic statement.

Benacerraf can be acquitted of merely confusing these notions only on the supposition that we have reason to expect a complete and unified semantic theory which includes a theory of truth as well as some component or other fulfilling each of the notions (1) through (4). In fact, though he never unequivocally spells it out this way, Benacerraf perhaps intends his semantical requirement on the philosophy of mathematics as the requirement not merely of a theory of truth, but of exactly that kind of comprehensive semantic theory.

Certainly Benacerraf is not alone in combining these notions. The very ease with which the passages mentioned glide by on a first reading measures the frequency with which these notions are conflated by philosophers of language. I believe that confusion is born of speaking at all of _the_ semantics of a particular expression or class of expressions. There are questions about expressions that are syntactic--that of the grammaticality of a particular expression, for example. There seem to be questions about expressions that are pragmatic, though these are somewhat controversial. Grice has claimed that the relation of implicature between sentences is a matter of pragmatics, and Goodman has claimed that only a pragmatic account of the projectibility of predicates can be expected to succeed. Equally there are questions about expressions that are semantic. Philosophers usually count questions of meaning, reference, analyticity, and entailment, as well as questions of truth, as all being semantic questions. I have no quarrel with this familiar classification, nor with the criteria that underlie it. With minor qualifications, I agree that issues about logical form as well as issues about the nature of mathematical truth are semantic issues insofar as they are issues in the philosophy of language; that is to say, they are semantic issues rather than syntactic or pragmatic issues. Perhaps I even agree that in some sense a completely satisfactory philosophical view would be required to include a semantic theory that settled all such issues, as well as answering all questions about meaning, reference, and entailment. But I am sure that this is a utopian requirement. It is a requirement only in the sense that nothing short of perfection can be completely satisfactory. No one has even begun to construct such a comprehensive semantic theory for any language of any complexity, whether for expressing mathematical propositions or others. And certainly no extant position in the philosophy of mathematics carries with it such a comprehensive semantic theory.

In Sections II through V, I explore the four notions listed above in order. In addressing each of them, it is important to keep in mind the questions whether the notion is one about which various philosophies of mathematics draw importantly different conclusions, and whether the attempt to construct the relevant account would be likely to lead a philosopher to a Platonistic conception of mathematics. In these sections, I argue that the several notions are conceptually distinct, that accounts of each type are, with respect to content, more or less independent of one another and independent of the theory of truth, that the notion of a statement of truth conditions is perhaps not philosophically significant, and that in any case no comprehensive semantic theory embracing all of these notions is to be expected on the basis of current achievements.

II. THE LOGICAL FORM OF ARITHMETIC SENTENCES

The question of logical form, in particular of the logical form of Benacerraf's sentence (2), surely cannot be the point of any interesting dispute in the philosophy of mathematics.[5] This will be the essential conclusion of the upcoming considerations, but keeping in mind our interest in disentangling various semantic notions, we want to survey the notion of logical form with some breadth. Along the way, I will defend these several other claims. Though an account of truth for many particular mathematical languages can be expected to begin on a basis that makes questions of logical form unproblematic, an account of logical form is distinct from an account of truth. An account of the logical form of the members of some set of sentences may also be distinct from an account of the semantics of various categories of subsentential expressions, like quantifiers or singular terms, though there are cases where it will not be. From my unequivocal tone, one might conclude that the notion of logical form is clear and unproblematic. Nothing could be further from the truth. It is not even clearly proper to speak of <u>the</u> logical form of (2),

[5] See p. 89 above.

or of any other sentence. Nevertheless, whatever is unclear or in dispute about the notion of logical form is not likely to affect its relationship to the other notions in which we are interested. Let us now investigate these claims in the reverse order of this summary.

Logical theorists invariably appeal to the notion of logical form with more assurance than that with which they explain it. At worst a definitive theory of logical form would require a solution of the most intractable problems of meaning and translation. At best it would depend on demarcating the scope of logic, that is, determining which inferences are licensed by logic alone and which truths are logical truths. In an unpublished paper Richard Cartwright has pointed out that neither the notion of logical form nor the more modest notion of difference in logical form have been elucidated by philosophers who use these notions.[6] More importantly, he has argued that elucidations of them utilizing properly modest logical devices are not likely to capture what we intuitively expect of the concept of logical form.

Insofar as we understand it at all, logical form is inferred from the inferences that we recognize as valid. The idea of logical form informs what is derivatively called formal logic. The idea is to develop a method of schematic representation for sentences such that combinatorial or otherwise algorithmic operations on these schematic representations will yield verdicts on such questions important to logic as validity, implication, and logical equivalence. A system of such schematic representations typically comprises what we usually call a formal system, what

[6] Richard Cartwright, "On the Concept of Logical Form," TS.

used to be called a logistical system. From the standpoint of any one of these formal systems, we would say that the logical form of a given sentence in English is exhibited by the formula that represents it schematically and that two sentences have the same logical form when they are schematically represented by the same formula. Of course no given sentence need be schematically represented by only one formula, even within a single formal system. After all, every sentence is an instance of the formula, 'P'. But the schematic resources of at least many formal systems are exhausted before schematization has gone very far, and within any such system we can say more precisely that two sentences have the same logical form when they share all of their schematic representations and that the logical form of a particular sentence is that formula, of all those of which it is an instance, that has the fewest instances. (Actually even this much does not isolate the logical form; witness '$(\exists x)Fx$' versus '$(\exists x)Gx$'.)

Two kinds of problem arise for this approach to logical form. Commentators agree that it is prior recognition of inferential relations that guides this kind of theorizing.[7] But this prior recognition does not fix the results of the theory. In the first place there may be crucial disagreements about the selection of some inferential relations as purely logical relations. And in the second place, even a fixed determination of the logical relations that the theory must reflect does not fix the

[7] See, for example, Gilbert Harman, "Logical Form," Foundations of Language, 9 (1972), 38-65; rpt. in The Logic of Grammar, ed. Donald Davidson and Gilbert Harman (Encino, Ca.: Dickenson, 1975), pp. 289-307.

assignment of logical form to the sentences under consideration.

Let us begin with the second problem. Suppose we agree about the purely logical relations that hold between members of an arbitrary set of sentences. How do we assign formulae to these sentences? Remember that our sketch used the notions of a formula schematically representing a sentence and, inversely, a sentence being an instance of a formula. What is the relation here being invoked? The deepest problem of philosophical analysis threatens to obstruct an explanation. Hoping to parry this threat, Cartwright assesses adequacy conditions that might be imposed on the relation, specifically considering these two:

(1) The (combinatorially expressible) logical relations on the schematic representations mirror those already determined to hold among the sentences at hand; and

(2) The truth values of the sentences at hand are reflected by some interpretation, in a model-theoretic or otherwise precisely specified sense, of the schematic representations.

Cartwright suggests that these adequacy conditions are not strong enough to exclude obviously unwanted candidates for the representation relation. To see what he has in mind, consider the set containing the sentence, 'Every number is even'; the sentence, 'Some number is even'; and the one thousand different sentences that can be constructed by replacing the variable in 'ν is even' by a decimal numeral from '1' to '1000'. What we want to do is to construct a perverse function from these sentences to some set of formulae that clearly are not the ones we want. Here is a schematization function that will do:

f('Every number is even') = 'P';

f($\ulcorner\nu$ is even$\urcorner$) = $\ulcorner P \vee T_{\nu}\urcorner$ or $\ulcorner P \vee F_{\nu}\urcorner$, depending on whether ν names an even number or an odd number;

f('Some number is even') = '$P \vee T_{1} \vee \ldots \vee T_{1000} \vee F_{1} \vee \ldots \vee F_{1000} \vee R$'.

Now we assign truth values to the sentence letters in those formulae, so as to reflect the actual truth values of the sentences:

I('P') = Falsity;

I($\ulcorner T_{\nu}\urcorner$) = Truth, for all values of ν;

I($\ulcorner F_{\nu}\urcorner$) = Falsity, for all values of ν;

I('R') = Truth.

This schematization function and interpretation together satisfy both of the adequacy conditions--(1) and (2), above--that Cartwright imposes on the relation of schematic representation.[8] Yet clearly these formulae from the propositional calculus do not schematically represent the

[8] It should be obvious how to tinker with the schematization function and interpretation to handle an arbitrary finite set of sentences constructed from a single predicate and a finite stock of singular terms by analogy with this set.

sentences we began with.

The insufficiency of the conditions, (1) and (2), leaves us wondering what further adequacy conditions could be imposed to isolate the relation of schematic representation. This is the central question that Cartwright calls to our attention. (He does not try to answer it.)

We saw that the standard approach to logical form uses a system of formulae to represent sentences schematically. Combinatorial manipulation of these formulae is expected to capture the inferential relations that actually hold among the sentences. Let us return now to the first problem that we raised for this approach, namely that there is room for disagreement about which inferential relations between the given sentences are purely logical. Differences over the prior classification of inferential relations arise even when there is no dispute about the actual truth values of the sentences under consideration and would remain in spite of agreement about which are their possible combinations of truth values. These differences reflect philosophical disagreement about the scope of logic.[9] Logical theorists disagree about how deep a schematization can go and still remain a logical schematization. Can the schematization capture second-order inferences, for example? And can it find structure within the boundaries of single words? These are well-known uncertainties. An interesting one suggested by remarks of Quine is similar.[10] We are quick to conclude that Mary is sibling to Martha,

[9]Cartwright is perfectly well aware of this, discussing it on p. 7.

[10] Willard Van Orman Quine, <u>Word and Object</u> (Cambridge: M.I.T. Press, 1960), p. 182.

given that Martha is sibling to Mary. Should our logical resources enable us to capture this inference? Should we have special schematic letters for symmetrical relations and a corresponding rule of inference: Given 'xSy', infer 'ySx'. Or is the symmetry of brother-or-sisterhood a nonlogical fact that should be registered in an explicit premise, with the schematic representative of the relation confined to the more usual rules of inference? Different verdicts yield different formal systems.

The practical response of logical theorists to these difficulties is modesty and relative tolerance. A variety of formal systems are propounded and employed. No one claims priority for one over all of the others in point of correctness, in contrast to ease of use, perspicuity, ease of inculcation, or economy of apparatus. Within a favored system, the student is taught logical schematization as a black art. And Quine piously hopes that schematization will "present little difficulty to one familiar with the canonical notation," inasmuch as he can "judge outright whether his ends are served by the paraphrase."[11] Even in our rough way we have an account of logical form only relative to a system of schematic representation.[12] The plurality of such systems are not isomorphic to

[11] Quine, Word and Object, p. 159.

[12] For some systems we would not have even that. What if we take up Quine's suggestion and add to it a similar device for transitive relations? Given the transitivity of (full) brother-or-sisterhood, two schematic representations would be possible for 'Mary is sibling to Martha', neither of them deeper or more basic than the other. (Actually the transitivity of that relation is not given, for though Mary be sibling to Martha and Martha be sibling to Mary, Mary is not sibling to herself. Another relation could be found to more accurately illustrate the point only at the price of artificiality.)

one another because their formulae cannot be correlated in the necessary way. For that reason the varying verdicts on the logical form of a particular sentence cannot be unified into one per sentence. Under pressure, speaking of the unique logical form of a sentence collapses into nonsense.

Even thus dimly understood, what relationship does this notion of logical form bear to the semantics of singular terms or of the quantifiers or of other categories of expression subordinate to the complete sentence? A theory of logical form may have no bearing whatever on the semantics of many such categories. For a case in point, notice that traditional treatments of first-order (nonmodal) logic assigned so-called individual constants to proper names without regard to questions about rigid designation and about causal chains of reference. These questions are now seen as central to the semantics of proper names, but the traditional treatments are no less viable for that. Since these semantic questions are irrelevant to the class of inferences being formalized, no answer to them is needed for the theory of logical form. And so in general a theory of logical form may take the meanings or other semantical properties of various types of expression as completely unproblematic and thus unexplained or, on the other hand, as awaiting further investigation. Except as these properties bear on inferential relations captured by the theory, it can leave them unexplained. In doing so it would assign to expressions in a particular category either a logical particle or a schematic representative (whichever may be appropriate to the expression at hand), with the result of licensing certain inferences involving sentences containing those expressions and barring others.

In certain cases though, an account of the semantics of a class of expressions is intertwined with a theory of the logical form of containing sentences. I have in mind cases where the expressions are given contextual definitions. The most familiar example is Russell's theory of descriptions. Here we give an explanation of the semantic role of definite descriptions by pairing every ordinary sentence that contains a definite description with a replacement incorporating the devices of quantification and identity and lacking the definite description. The logical form of the replacement is manifest. Precisely because the inferential relations borne by sentences containing definite descriptions are complex and controversial, accounts of logical form, which predict and account for these inferential relations, delimit what can be said of their semantical properties. But this interdependency should not be exaggerated. Precisely because the definition thus given is a contextual definition, it bypasses important questions about the meaning of definite descriptions. How is it that the various morphemes comprising the description when combined in this way with the rest of the sentence yield a whole that has the logical relations that it does? Similarly the account of logical form can proceed without defining denotation for definite descriptions, though other purposes might make defining this semantic relation worthwhile. In no wise are we justified in calling Russell's theory of descriptions a complete semantical theory for definite descriptions.

What of the relationship between accounts of logical form and theories of truth? For a modest enough language, truth could be defined without paying attention to issues of logical form. So long as the language be finite, no matter how large, there is a theory of truth for it that ignores the syntactic structure of its sentences. But in interesting cases, recursive rules of formation generate the infinitude of sentences in the language. Where these syntactic rules of formation are well-behaved, an inductive truth definition can be specified, mimicking the structure of the formation rules. Only formal languages are well-behaved in this way, and of infinite languages only formal languages have ever been provided truth definitions. Of course the sentences of these languages wear their logical forms on their sleeves. For that reason, I said that when a truth definition can be given, logical form is unproblematic. Nonetheless, the definition of truth is not the same as the account of logical form. The logical form of the sentences of a formal language is implicit in their syntax. The truth definition only begins with a specification of syntax.

Without at least a material correlation of the sentences of a language whose syntax was more complex to the sentences of a formal language, no truth definition could be given for it. Would such a correlation fulfill any conceivable adequacy conditions for a theory of logical form and thus implicitly be a theory of logical form? Or would the difficulties Cartwright finds with plausible adequacy conditions haunt this case as well, leaving us in the dark about logical form even though we had an extensionally correct truth definition? We cannot stay for an answer.

What is more important for us is that none of this uncertainty about the notion of logical form and none of the particular disputes about the best system of schematizing this or that set of sentences correlate to uncertainties in or disputes about the philosophy of mathematics. Arithmetic is the home of the modern analysis of first-order inference. All of the early formal systems were developed with an eye to systematizing the almost universally accepted inferential practice of mathematicians. There is one important exception to this universal acceptance, namely, intuitionism. Formal systems have their motivation and their justification in capturing standard classifications of inferences. It is well-known that intuitionists dissent from important examples of standard classifications. But they do not express this specific dissent in terms of logical form. Rather than rejecting particular schematizations of various sentences, they reject particular syntactic rules that certify inferences from several such sentences to a conclusion. For example, they accept the usual schematization of two sentences as '~~P' and 'P', but reject the rule licensing the inference of the second from the first.

Having said this much, we should return to Benacerraf's worries about the logical form of (2), his typical arithmetic sentence.[13] It seems clear that even when we have settled on the logical form of (2), we must go further to find out what it means to say that (2) is true, and that we will have hardly begun to explain the semantics of its component expressions. And it seems clear that, given that we accept the system of schematization exemplified by (3), and given that we use (3) to represent the geographical sentence, (1), we would be silly to deny that (3) is the representative of (2) as well. Presumably the logical relations entered into by (2) are accounted for by means of (3), and what more do we intend in thinking of (3) as giving the logical form

[13] See p. 89, above.

of (2)?

III. THE SEMANTICS OF SINGULAR TERMS IN ARITHMETIC

Seeing Benacerraf inquire into the semantics of numerals and other singular terms in the language of arithmetic, we most naturally wonder about the relationship between theories of reference and theories of truth. After all, Benacerraf sets up his semantical requirement as if he expects there to be important relationships between the various semantic accounts, and the notion of reference is perhaps the most trusted of the semantical properties of singular terms.

The relationship between truth and reference and between theories of each comprises the subject of "Field's Theory of Truth."[14] What are the relevant conclusions there formulated? One important conclusion for present purposes is that only if a language is particularly simple can a theory of truth be intelligibly based on a primitive notion of denotation. Its singular terms must be limited to a so-called semantic category that is rather clearly characterized by at least a rough criterion of denotation for its elements. In asking for a unified account of the reference of both numerals and singular terms that denote other sorts of things, Benacerraf runs afoul of an important point made in that essay, namely, that reference mediates so many different relata of so many different kinds that we have a fix on what relation it is supposed to be only by

[14] That is, the second part of this thesis, beginning p. 47, above.

way of the disquotational paradigms: 'Socrates' refers to Socrates; 'π' refers to π; and so on.[15]

Equally importantly, we saw in "Field's Theory of Truth" that a definition of truth that is completely independent of any prior notion of reference can be given for any static language. The original theory of Tarski, there paraphrased as T2, is an example of such a theory. Of course, as we noted, an extensionally correct definition of denotation can easily be formulated in the process of supplying that kind of theory of truth.[16] But such a definition of denotation will be unavoidably parochial, bound to the vocabulary of the language under consideration in its current state. It will hardly constitute a general theory of denotation, and it will neither issue from nor contribute to a comprehensive semantic theory.

Other aspects of the semantics of singular terms, for example, their meanings, their designation across possible worlds, the circumstances historically responsible for their referring to what they do or for their meaning what they do, seem to be completely independent of the concept of truth and of the theory of truth. And in considering logical form, we saw that semantical properties of singular terms may or may not be independent of a theory of logical form, depending on what singular terms are involved. Contextual definitions of singular terms, as in the case of Russell's theory of descriptions, implicate matters of logical form; straight definitions do not.

[15] See pp. 74-76, above.

[16] See p. 59, above.

With regard to arithmetic singular terms in particular, the most pressing semantical issues have to do with numerals. Benacerraf raises the question whether numerals are names. They do not seem to be proper names. Even in the absence of a characterization of proper names, there seems to be a clear difference between proper names like 'Socrates', 'London', and 'Harvard', and numerals like '1' and '456'. In fact a case could be made for regarding standard numerical designators beyond '9' as definite descriptions. On the other hand, numerals no doubt are rigid designators, inasmuch as arithmetic consists entirely of necessary truths. It seems that neither the question whether they are proper names nor the question whether they are rigid designators can be what Benacerraf wonders about.

Perhaps Benacerraf means to be asking whether there are any things to which numerals bear the denotation relation. My reply would be that the standard definition of denotation (for example, (D) on p. 59, above) certainly makes it look like there is. For example, it looks at first appearance like '12' denotes 12. But similarly it looks at first appearance like the Apostles are numbered by 12. And yet philosophers have called into question whether there is anything which the Apostles are numbered by. Since no one wants to deny that we can use existential generalization over numerals (Do they?), whatever is at stake depends on the ontological force of the existential quantifier. This is an issue that we will defer to our discussions of Platonism, in Section VII.

Rather than the dilemma that he does pose for the philosophy of mathematics, Benacerraf might well have posed another, this one ignoring epistemological matters. Because he is tempted to expect that the standard semantics will base its theory of truth on a theory of reference, I can imagine that Benacerraf would be sympathetic to what has been called the causal theory of reference. But the causal theory of reference seems itself to be incompatible with Platonism, understood as the doctrine that numbers are abstract entities whose existence is independent of human thought. In light of difficulties about the explanatory status of a causal theory of reference, we might just look at a minimal causal theory, one that is only extensional: Is there a causal relation extensionally equivalent to the denotation relation for numerals? Or more modestly, even for the numerals from '0' through '9'? I suppose that there is not, given that the numbers are abstract and do not enter into causal relations. And so much the worse for the causal theory.[17] In any case, though this mathematical issue grows out of considerations about the semantics of numerals, it is a parochial one that only arises on a particularly strong conception of the standard semantics for mathematics.

In summary, I have argued that, despite Benacerraf's hopes, there is no reason to expect that a theory of truth will be founded upon a theory of reference in an interesting way, that doubts about whether numerals denote numbers are no different from other doubts about the reality of numbers and need be handled in no different way, and that the causal theory of reference is incompatible with the ordinary characterization of

[17] I suspect that this dilemma, as much as Benacerraf's official dilemma, motivates Field's nominalism in Science without Numbers. Of course there have been attempts to locate numbers in the causal relations that are supposed to underlie knowledge; I think of Steiner. These same efforts might be effective against my more narrow dilemma.

numbers as abstract objects. I can find no other issue in the philosophy of mathematics that has much to do with the semantics of numerals and other singular terms in arithmetic. Foundational positions in the philosophy of mathematics did not speak to the question of the semantics of numerals or other singular terms, except wholeheartedly to embrace Russell's theory of descriptions as a part of the formalization of mathematical reasoning. Logicism and formalism both accepted mathematical formalization; intuitionism rejected the whole package, the theory of descriptions included. Now let us turn to questions about the semantics of quantifiers in arithmetic, asking whether these questions significantly relate to the theory of truth.

IV. THE SEMANTICS OF QUANTIFIERS

It is difficult to guess what Benacerraf has in mind when he speaks of a semantic account of the quantifiers in arithmetic. What seems to be called for is a definition of the quantifiers as they appear in the language of arithmetic, but obviously the kind of definition appropriate for a term or a relation, or even for a connective, will not do. Quantifiers are operators that (when successively applied) turn open sentences lacking truth values into closed sentences having truth values. They are inseparable from the variables or other pronominal devices that they bind. Given this, perhaps they could be explained on the model of the semantical explanation of definite descriptions. Contextual definition seems the only kind of definition possible for an operator. As such the definition would, for any sentence in which quantifiers occur, generate an equivalent in which no quantifiers occur, and would do so by some systematic means. So far as I know there is no correct account of the quantifiers on this construal, either for the language of arithmetic or for any other language that is powerful enough to describe an infinite domain. Only a few have thought this kind of definition of the quantifiers possible--I think of Wittgenstein in the Tractatus--and for that reason only a few have even tried to construct one.

Nevertheless, questions about the nature of the quantifiers have been at the center of important philosophical disputes in the century since Frege first enunciated a logical theory that employed quantifiers and systematized quantificational inference. The philosophy of mathematics has been a particularly frequent scene of these disputes.[18] Further discussion of the quantifiers and the philosophy of mathematics, focused on the question of the ontological force of the existential quantifier, appears in my treatment of Platonism. There we will be asking whether mathematical objects can be explained away, while the statements of mathematics retain their truth values. (See pp. 156-158, below.)

Another kind of semantic account has often been thought relevant to these disputes. I refer to the familiar proceedings which are usually described as specifying the semantics of a formal system and which include elements devoted to the quantifiers. These proceedings occur in almost any serious exposition of formal logic; a particularly compact and easily

[18] For a thought provoking account of some little known aspects of these disputes, see Warren D. Goldfarb, "Logic in the Twenties: The Nature of the Quantifier," The Journal of Symbolic Logic, 44 (1979), 351-358.

identified instance occurs in Mates' Elementary Logic.[19] The point of this exercise is to provide an interpretation or a number of interpretations for the formal system. The purpose of providing interpretations in the technical sense exemplified, could be either one of two:

(1) Converting a meaningless formal system into a meaningful formal language, or

(2) Artificially assigning truth values (under the interpretation) to formulae in a way that is systematically related to the possible truth values of the ordinary statements that they represent.

If the formal system is newly introduced and so far only syntactically determined, the process of supplying an interpretation gives life to an heretofore flat and meaningless system of symbols. For those who think specifying conditions of the truth of a statement adequate to render it fully meaningful, an interpretation is seen to convert the meaningless formulae of a formal system into sentences of a fully meaningful formal language, thus serving the first purpose. Of course this requires that the interpretation span the whole formal system, as it typically does.

The second purpose, that of artificially assigning truth values, is served even for those who deny that fixing truth conditions is sufficient to give meaning to an otherwise meaningless symbolism. Explaining what

[19] Benson Mates, Elementary Logic, 2nd ed. (New York: Oxford University Press, 1972), p. 60.

an interpretation is, and doing it in such a way as to make evident what are the range of possible interpretations, gives content to the standard definitions of validity, implication, and logical equivalence for formulae of a formal system, definitions expressed in terms of possible interpretations of the formulae. The cogency of these definitions depends not on the assumption that the formulae so intepreted themselves become fully meaningful sentences, but only on another assumption that is usually unstated. The crucial assumption is that for each statement of the informal language of ordinary discourse, we can be sure that the formula which schematically represents it has the same truth value as the ordinary statement, under some interpretation which is systematically related to the ordinary statement. Part of the systematic relation, for example, is that the class assigned by the interpretation to a predicate letter be the extension of the corresponding predicate of the ordinary sentence.

Two kinds of misconception easily develop about such semantical accounts of formal systems. The first is that to provide a semantics in this way is to provide a definition of truth of the sort that Tarski provided in his well-known theory of truth. Though he no doubt did not suffer this misconception himself, Mates introduced the semantics for his formal system using a phrasing that may be one source of this confusion. The relevant section of his text is entitled "Truth," and he begins,

> Next we wish to make clear what it means to say that a sentence of L is true or false with respect to a given interpretation of L. In other words, we seek to give an exact definition of the locution
>
> ϕ is true under I,

> where the values of 'ϕ' are sentences of L and the values of 'I' are intepretations of L.[21]

The promise to define "ϕ is true under I" can easily be misconstrued as a promise to define truth itself. The misconstrual is reinforced by the fact that the definition of truth under an interpretation is structurally very similar to Tarski's definition of truth itself, or more accurately, of truth in a language.

Against this first misconception, I contend that a specification of formal semantics does not serve as an explanatory definition of the concept of truth. A simple argument for this contention is the observance that truth under an interpretation is not the same thing as truth. Yet there is a view that this mere observation will not defeat. I have in mind the view that the formal semantics does effectively define truth for the formal languages, but only when what we have seen so far is supplemented by choosing one of the intepretations as the intended one. Then the formulae would be true or false as they were true or false under that special interpretation. My reply to this view depends on reminding ourselves of the two possible purposes of formal semantics outlined above, namely, either giving meaning to the formulae or else artificially assigning truth values to the formulae under various interpretations, which systematically mirror the various possible truth values of the corresponding ordinary sentences. Giving meaning to an otherwise meaningless formula by providing truth conditions for it depends for its efficacy on a prior understanding of what truth is. Similarly, mirroring the truth

[21] Mates, Elementary Logic, p. 57.

values of ordinary sentences by assigning truth values to a schematic representation under an interpretation that is systematically related to the structure of the ordinary sentence is of no value in the absence of an understanding of what truth and falsity regarding the statements of the ordinary language are all about. The point of Tarski's definition of the concept of truth, and, I suppose, of any other philosophical elucidation of truth, is this prior one of providing the background understanding of truth.

The second common misconception about semantical accounts of formal systems is that those clauses of such accounts that are devoted to the quantifiers constitute an explanation of the semantics of quantifiers generally. Those who think of truth conditions as sufficient for the explanation of meaning are particularly susceptible to this misconception, inasmuch as the clauses in question do provide truth conditions for quantified formulae of the formal system. Let us take that clause in Mates' formal semantics which is devoted to the existential quantifier as an illustration. It says,

> If $\phi = (\exists\alpha)\psi$, then ϕ is true under I if and only if $\psi\ \alpha/\beta$ is true under at least one β-variant of I.[22]

Now I grant that if providing truth conditions is sufficient for conferring meaning, and if the point of the formal semantics is to confer meaning on an otherwise meaningless formalism, then this clause does confer

[22] Mates, Elementary Logic, p. 60. My point does not depend on an understanding of Mates' notions expressed by the notations, "$\psi\ \alpha/\beta$" and "β-variant".

meaning on those formulae that are existential quantifications. But its doing so depends on a prior understanding of existential quantification in another language, namely the language in which the formal semantics are stated, which usually is a natural language. The obvious reason is that the clause quoted makes use of such an existential quantification, articulated in the words "at least one β-variant". And so the clause does not serve as an explanation of existential quantification in general. On the weaker presupposition that meaning is not conferred by the specification of truth conditions, or that the formal semantics is not intended to make formulae meaningful anyway, but merely to define an otherwise useful notion of truth under an interpretation, there is no justification whatever for the idea that the existential quantifier clause contributes to a general explanation of the semantics of the quantifiers.

Since surveying Benacerraf's dilemma in Section I, we have considered the notions of a theory of logical form, of an account of the semantics of numerals and other arithmetic singular terms, and of an account of the quantifiers. We have distinguished all of them from the notion of a theory of truth and assessed their own prospects along the way. The remaining notion on our list, one which Benacerraf often employs in expounding his dilemma, is that of a formulation of truth conditions of an arbitrary arithmetic statement. In this section, we have seen a dubious appeal to the notion of truth conditions in connection with endowment of meaning, so it is appropriate now to turn to that notion.

V. THE CALL FOR TRUTH CONDITIONS

In many passages Benacerraf asks after the truth conditions of typical arithmetical statements. His doing so calls to mind the frequent appeal to the notion of truth conditions in contemporary philosophical inquiries, particularly in inquiries into the theory of meaning, and even more particularly in inquiries into the meaning of mathematical statements. What is it that Benacerraf is asking after? What are truth conditions, anyway?

Our first temptation is to think that by "truth conditions" must be meant necessary and sufficient conditions for truth. This supposition fits Benacerraf's inquiry into a long tradition according to which philosophical insight is achieved through discovering necessary and sufficient conditions for the application of some central concept. Unfortunately the notion of necessary and sufficient conditions itself conceals an important distinction. Sticking with the case in point, when we say that some condition is necessary and sufficient for truth, we may mean merely that if the condition holds for a statement, then the statement is true, and conversely, if a statement is true, then the condition holds for it. So to give necessary and sufficient conditions for truth is to supply an extensionally correct definition of truth. To insure that the definition is not trivial by reason of circularity, we usually require that the condition not itself make use of the notion of truth. Departing from this weak sense of necessary and sufficient conditions, many philosophers find mere coextensiveness inadequate for their purposes in seeking conditions for a concept. For them the _necessity_ of necessary conditions is to be taken seriously. A genuine definition of a concept, they say, does not merely specify those things which happen to fall under the concept; it must capture the meaning of the concept and thereby provide an explanation why these things and no others do fall under the concept. These philosophers seek, under the description "necessary and sufficient conditions", what has in this century been called a philosophical analysis of the concept and thereby clearly commit themselves to the troubling distinction between analytic and synthetic statements.

Having distinguished this stronger notion of necessary and sufficient conditions, we will be aided in the ensuing discussion by noting some candidates for necessary and sufficient conditions for truth in at least the weaker sense of extensional equivalence. There follow several examples.

(1) It may well be a necessary and sufficient condition for the truth of any statement that God believe it. If you suspect the condition of covertly appealing to the concept of truth itself, take God to be a proper name devoid of meaning, or take it to mean the creator of the world, rather than taking it to mean an omnipotent, _omniscient_ being.

(2) In the case of a syntactically complete mathematical theory, it is a necessary and sufficient condition for the truth of any statement in the theory that it be a logical consequence of the axioms.

(3) In the case of a language consisting of all the simple Noun Phrase-Verb Phrase sentences generated from the noun phrases 'roses' and 'violets' and the verb phrases 'are red' and 'are blue', it is a necessary and sufficient condition for the truth of any statement expressible in the language that it be expressed either by 'roses are red', or by 'violets are blue'.

(4) It is a necessary and sufficient condition for the truth of any statement expressible in the language specified in (3) that one of the following hold:

(a) 'Roses are red' expresses the statement in question, and what's more, roses are red;

(b) 'Roses are blue' expresses the statement, and roses are blue;

(c) 'Violets are red' expresses the statement, and violets are red;

(d) 'Violets are blue' expresses the statement, and violets are blue.

(5) It is a necessary and sufficient condition for the truth of any statement expressible in the language of (3) that, with reference to the sentence expressing it, the entities denoted by the noun phrase satisfy the verb phrase.

It should be amply evident that formulating necessary and sufficient conditions for the truth of statements expressible in some language or other is not a very difficult task. As is illustrated by (3), (4), and

(5), there need not be only one such set of conditions. A further illustration would be provided by an alternative axiomatization of the theory in (2). And nothing precludes conditions for the theory in (2) of the sorts illustrated by (4) and (5), or at least by a familiar extrapolation of them to handle infinite languages. Of course, if it turns out that (1) successfully mentions an omniscient being, it provides truth conditions for all statements irrespective of the expressive power of particular languages.

In giving a philosophical account of truth one must face the central problem of distinguishing among the many sets of necessary and sufficient conditions, in at least the weaker sense, that could be formulated. A theory of truth presumably will provide necessary and sufficient conditions, but in order to be philosophically illuminating, it must do more than this. What more can reasonably be required of a definition of truth than mere extensional correctness in a question to be asked (and answered) in our next section.[23] But let us note now that there is no reason to expect the requirements to lead to what would traditionally have been called an analysis of the concept of truth, namely _analytically_ necessary and sufficient conditions. By way of confirmation, we should notice that no one of our five examples qualifies as an analysis.

Benacerraf often speaks of _the_ truth conditions of mathematical statements, as if there were only one set of necessary and sufficient conditions common and peculiar to mathematical statements. In other passages though (for example, p. 666), he speaks of _a_ truth condition,

[23] See pp. 127-128, below.

apparently recognizing what we have emphasized, that there are many sets of conditions for truth. However we understand him on this score, we must guard against one suggestion we may be left with. In introducing his epistemological requirement, Benacerraf says, "the conditions of the truth of mathematical propositions cannot make it impossible for us to know that they are satisfied" (p. 667). We might wonder whether, if (1) turns out to hold good, the knowability of mathematical propositions is not threatened. If the condition of the truth of some proposition is that God believes it, isn't it impossible for us to know that this condition is satisfied? After all, isn't God inscrutable to man? Of course the proper reply will be that one set of conditions may seem inacessible to mere mortals, while another is quite accessible. (Of course if the other is accessible, the first will turn out, contrary to appearances, to be accessible too, though indirectly, supposing of course that they both really are known to be necessary and sufficient conditions for truth.) Truth conditions of the type of (1) or even of the type of (5) may seem quite impossible of general determination, while those of type (2) would be quite easily determined in many cases. So long as conditions of the type of (4) are included among the alternatives, there should never be any general obstacle to knowledge growing out of conditions on truth.

When Benacerraf asks after the truth conditions of particular arithmethic statements, as we saw him do earlier, he seems to expect a different answer for each statement.[24] The idea seems to be that each particular statement has its own truth conditions, distinct from those of other

[24] See p. 89, above.

statements, even other statements expressible in the same language. This idea may not be borne out on the characterization of truth conditions that we have so far considered. If (1) does not fail of reference, and if God indeed is omniscient, then the condition for the truth of every statement will be that God believe it. This necessary and sufficient condition for the truth of any statement whatever makes use of no peculiar properties of given statements; neither structural, inferential, semantical, nor any other peculiar properties are exploited in the truth condition. For that reason the general statement of truth conditions in (1) does not entail distinct truth conditions for distinct statements. Similarly, from the general condition of derivability from a set of axioms set out in (2), there follow no particular derivability properties for the distinct true statements of the mathematical theory. Yet in interchanging the notion of the truth conditions of a statement with that of its logical form, and in expecting the truth conditions to be determined by or importantly related to the semantical properties of the expressions that make up the sentence, Benacerraf takes it for granted that the truth conditions will be particularized. In order to maintain contact with the traditional notion of general necessary and sufficient conditions, in this case conditions on truth, he must expect the general condition to entail distinct conditions on the truth of distinct statements. Benacerraf has this expectation in common, I believe, with a prominent strain of contemporary philosophical discussion that makes heavy use of a particularized conception of truth conditions. A fact complicating our discussion is that this strain is devoted primarily to investigating and generalizing issues in the philosophy of mathematics. I have in mind the

the explorations of contemporary philosophers of language into the theory of meaning, and the general contrast that they draw between truth-conditional and verificationist theories of meaning.

An analysis and assessment of this strain of discussion would take us far outside the focus of this study. I mention it only in order to acknowledge the relevance to issues in the philosophy of mathematics of the notion of the truth conditions of a statement, the relevance in particular of the question whether an explanation of meaning can be given in terms of truth conditions. I have no doubt that Benacerraf does not have this question in mind when he asks after the truth conditions of particular arithmetical statements. Nevertheless, there is something to be learned from considering truth-conditional accounts of meaning, something regarding the possibility of formulating philosophically interesting truth conditions for individual arithmetic statements. Though Benacerraf never calls for a truth-conditional account of meaning, he does seem to think that _something_ would be explained by the particular truth conditions of arithmetic statements.

The truth-table account of a sentential connective is a paradigm case of a truth-conditional specification of meaning. What a truth table for the material conditional tells us, to consider an example, is that a conditional sentence is true if the antecedent and consequent are both true, or the antecedent is false and the consequent is true, or the antededent and consequent are both false, and that the sentence is false if the antecedent is true and the consequent is false. This is not itself an account of truth conditions peculiar to an individual sentence, but it easily results in one: 'if roses are red, then violets are blue' is true if 'roses are red' and 'violets are blue' are both true, and so on. What is instructive about this example is that the conditions of truth specified for the particular sentence are informative, even for someone who does not understand the sentence or who has never before seen the material conditional. Specifically, the truth conditions are not expressed by means of the sentence under consideration. Furthermore, some simplification has occurred. Whereas the conditional sentence might be said to describe a complex state of affairs, the truth-conditional account of it is given by means of descriptions of simpler states of affairs. (Perhaps this claim can be supported only if the truth conditions are reformulated using semantic descent, 'roses are red' replacing '"roses are red" is true' in the statement of truth conditions, and similarly with the other clauses.)

Unfortunately, there is no reason to expect to be able to imitate, with respect to arbitrary sentences, the truth-conditional account of the meanings of complex truth-functional sentences. Only on the basis of metaphysical or linguistic doctrines that are not widely held could we expect a truth-conditional account of every given sentence that was both systematic and informative. One example of a philosopher who could seriously inquire into the truth conditions of particular propositions was Wittgenstein in the period of the <u>Tractatus</u>. According to the doctrine of that work, an informative and illuminating response can be given to the question, "What are its truth conditions?" asked of an arbitrary proposition, but only by reason of, (1) the truth-functional nature of <u>all</u> complex propositions, (2) the existence of states of affairs, and (3) the picture theory of elementary propositions.

Bereft of these presuppositions, or others equally problematic, as I suppose Benacerraf is, along with most of the rest of us, we have no reason to expect that anything interesting can be said in response to the question, "What are its truth conditions?" As we have seen, not all general conditions on truth issue in distinct conditions for distinct statements. Of the illustrative truth conditions that we set out earlier, only (4) and (5) issue in a peculiar set of necessary and sufficient conditions for each statement of the language. The general condition in (4), when supplemented by the obvious principles of identity, yields the particular condition for the truth of "roses are red" that roses be red, and similarly for the other sentences of that small language. Surely Benacerraf does not achieve philosophical satisfaction in the realization that the conditions of the truth of 'there are at least three perfect numbers greater than 17' is that there be at least three perfect numbers greater than 17.

There is one outstanding trump card against which we have had to finesse. The general set of truth conditions listed in (5) also yields particular truth conditions that may seem to be of philosophical interest. I suspect that Benacerraf would insist that it is an achievement to learn that a condition of the truth of 'roses are red' is that the referent of the noun phrase 'roses' satisfies the verb phrase 'are red', or to use another phrasing, that there are objects that 'roses' denotes and that 'are red' applies to them. And he would be correct if only it were possible to say something interesting both about the reference relation and about either the relations of satisfaction or its converse, applying to. Benacerraf recognizes this burden, acknowledging that his favored general condition on truth "must proceed through reference and satisfaction and, furthermore, _must be supplemented with an account of reference itself_" (p. 677, emphasis added). I doubt that it is possible to give an interesting account of the relations of reference and satisfaction, but this is not a place again to plead my case. The best explanation of these relations that I expect philosophers to be able to devise leads these apparently interesting truth conditions to collapse into those dismissed last paragraph. The only theoretically grounded condition of the truth of 'roses are red' in particular will, I believe, turn out to be just what that sentence says, namely, that roses are red.

Something more cries out to be heard. In recent years the truth conditions of various particular propositions or categories of proposition have been sought more and more often in philosophical contexts which make it clear that the real quarry is what in former years would simply have been called the _analysis_, not of truth, but of the proposition or type of proposition under consideration. The genesis of this manner of speaking is entirely innocent; if I seek an analysis of p, I seek analytically necessary and sufficient conditions for p, that is, analytically necessary and sufficient conditions for the truth of the proposition that p. That is, I seek the truth conditions of p. However innocent its origin, the new manner of speaking has the entirely specious advantages of sounding more scientific than the old, of not calling to mind decades of philosophical agonizing over the nature of analysis, of falsely insinuating itself into the domain of the theory of reference, and of not associating itself with the widely discredited analytic-synthetic distinction. There are real advantages in sticking with the old terminology, including the obverse of each of these counterfeit ones. We all know what is at stake in the call for a philosophical analysis. Our agony over the nature of philosophical analysis has not been for nothing, and we should not lightly cast off its fruits. Even philosophical failure results in an enlargement of philosophical understanding.

VI. THE THEORY OF TRUTH FOR MATHEMATICAL STATEMENTS

Over the course of the last five sections I have distinguished the notion of a theory of truth from notions of other kinds of semantical accounts, including accounts of logical form and accounts of important semantical properties of singular terms and quantifiers. I have also disparaged the notion of an arbitrary statement of the truth conditions of a proposition. Not only have I maintained conceptual distinctions among these notions, but I have argued that no comprehensive semantic theory for the language of mathematics (or for any other language of any complexity) that includes components answering to each of these notions will be developed. In fact I have urged that none of these kinds of accounts forms any very important part of any of the others. Through all of this I have said very little about what I take a theory of truth to be. Yet, I have sought to put the notion of a theory of truth at center stage. In fact, I have said that the most intelligible rendering of Benacerraf's semantical condition is that a satisfactory philosophy of mathematics must include an account of truth for the sentences in which mathematical propositions are expressed. It is time now to turn to the notion of a theory of truth. In this section I first explain what I mean by a theoretical definition of truth, in order to distinguish it from an arbitrary condition on truth. Then I defend the theoretical definition of truth formulated by Tarski, as fulfilling Benacerraf's requirement of an account of truth for mathematics.

A. Theoretical Definitions

When Benacerraf speaks of an "account" of truth, I take him to mean a philosophical elucidation or analysis of the concept of truth.[25] I believe that the preferable form of elucidation of a concept in philosophy, as in other forms of discourse, is a definition that is part of a larger theory. When I speak of a definition, I mean a theoretical proposition of the form,

> For any sentence s expressing a proposition of mathematics, s is true if and only if $\phi(s)$,

where '$\phi(s)$' stands for a sentence of some language in which the theoretical claim is expressed, open only in the variable 's'. For the sake of generality, we should embellish the form of the theoretical definition to provide for falsity as well, giving,

> For any sentence s expressing a proposition of mathematics, s is true if and only if $\phi(s)$, and s is false if and only if $\psi(s)$.

[25] Yet I am not confident in attributing to Benacerraf any particular view about so controversial a topic as what a philosophical elucidation consists in. I expound a view of my own in what follows without finding any definite suggestion of it in Benacerraf's article.

Many would presume without further ado that what replaces '$\psi(s)$' should simply be the negation of what replaces '$\phi(s)$', but we need not so commit ourselves.

In supposing that this definition constitutes a theoretical claim, I suppose that from the truth definition, perhaps together with some other theoretical propositions, there will follow important generalizations that we all accept, involving the notion of truth, as well as generalizations of interest that we have not otherwise been able to justify to our satisfaction. Generalizations of the first sort might well include what Dummett calls the Principle of Exclusion, that is, the principle that no sentence (here, of mathematics) is both true and false,[26] as well as other uncontroversial laws of logic or semantics. Examples of the second sort of generalization are the (semantic) completeness of a particular system of formal derivability or more controversial principles of logic or semantics. In short, in order to qualify as a theoretical definition, the truth definition must be theoretically productive and theoretically interesting.

I also suppose that the truth definition, in order to be a theoretical claim, will have an epistemological status characteristic of theoretical propositions. It must be well-supported by evidence or other epistemologically justifying features. Typically, theoretical propositions are justified by reference to their consequences. A definition

[26] Michael Dummett, Truth and Other Enigmas (Cambridge: Harvard University Press, 1978), p. xix. This is the semantic principle corresponding to the Law of Contradiction.

that yielded widely accepted semantic generalizations, as lately outlined, would thereby grow in plausibility. But consequences involving truth of specific form, if obvious and intuitively central to the concept of truth, would equally contribute to the justification of the definition. By placing a premium on consequences that are intuitively bound up with the very notion of truth, we avoid depending on consequences, however obvious, whose evidence rests on entirely extraneous considerations, for example, that God knows everything or that two epistemologically unrelated propositions have the same truth value. If there are other, perhaps more elusive, features characteristic of epistemologically well-supported theories--I think of elegance, simplicity, initial plausibility--the truth definition must, in order to qualify as a theoretical definition, exhibit them as well.

I believe that it is this theoretical character that would distinguish a definition of truth from other generalizations of the same form that, as we would ordinarily say, merely happen to hold. It need have no other special epistemological character such as _a priority_. And it need have no special metaphysical or semantic status such as necessity or analyticity.

Perhaps we should linger a moment on this question. In defining a concept, or classically in providing a philosophical analysis of it, one has in many contexts been expected to explicate its meaning. In doing so one would supply an equivalent that would preserve truth even in intensional contexts. I make no pretense that the kind of elucidative definition I am defending has this character. Indeed the paradox of analysis shows that this expectation is absurd for ordinary belief contexts.

A better model is provided by scientific definitions; no one supposes that the theoretical definitions of scientific terms like 'kinetic energy' or 'entropy' preserve truth under substitution in belief contexts. Nevertheless, if we are to be seen as giving a philosophical elucidation, not merely (!) a scientific theory, shouldn't something stronger than extensional equivalence be required? Of course I have required something stronger than extensional equivalence, but is what I have required enough? Shouldn't truth be preserved at least in modal contexts for the account to qualify as a philosophical account? Well, whatever you think of this requirement in general, there is some interest in seeing why it quite possibly cannot be fulfilled in this case.

Truth values are not likely to be preserved under substitutions of _definiens_ of the sort we have considered into modal contexts. This is not a result of some deficiency in the truth definition as such, but rather is a result of proceeding by defining truth for sentences in the first place. In order for modal substitutions to preserve truth, this stronger version of the definition must obtain:

> Necessarily, for any sentence s expressing a proposition of mathematics, s is true if and only if $\phi(s)$, and s is false if and only if $\psi(s)$.

The difficulty immediately suggests itself that a sentence might under different circumstances have a different meaning than it in fact has. Thus, if, contrary to fact, '5 + 7 = 12' meant that four times three is greater than seventeen, the proferred account of truth would very possibly fail. Some have replied to this worry by observing that in specifying a sentence of a language we specify not merely an alphabetic or phonetic string, not merely a syntactic object, but we specify an interpretation as well. On this view every expression that is part of a language has its meaning essentially. However plausible this may be, it is not quite enough to guarantee the stronger (modal) version of the truth definition. The reason is that on some views fixing the meaning of a sentence may not be enough to fix its truth value, because it may not be enough to fix the proposition that the sentence expresses. (Whether or not he intended them this way, a particular way of looking at Putnam's examples, "Water is H_2O" and "Cats are animals", will yield the point I want here.[27]) Setting this worry to rest would require an even stronger claim, namely, that every sentence expresses the proposition it does essentially. I see no way to argue for this claim; in fact it seems quite implausible to me. Many sentences express different propositions from occasion to occasion of utterance. It is true that we can specify languages whose sentences are complete (or eternal) in the sense that a sentence in fact expresses the same proposition on every occasion of utterance. But the task of specifying even a fully meaningful language whose sentences express the same propositions despite however great variations of use and social arrangement and scientific fact I do not know how even to begin. Furthermore, even if it did turn out that for some languages, each sentence expresses the proposition it does essentially,

[27] See Hilary Putnam, "It Ain't Necessarily So," in his Mathematics, Matter and Method (Cambridge: Cambridge University Press, 1975); and "The Meaning of 'Meaning'," in his Mind, Language and Reality (Cambridge: Cambridge University Press, 1975).

we have no guarantee that among them would be languages in which the propositions of mathematics could be expressed.

A definition of truth for sentences in which mathematical propositions are expressed could avoid this entire problem in only one way. The definiens would have to be complex in a way that enabled its verdict on truth value to vary in concert with all possible variations in the propositions expressed by sentences in the domain of the definition. In effect this would require the definition of truth for sentences to include an account of the relation between a sentence and the proposition it expresses. That is to say, a definition of the relation of expressing would have to be embedded in the definition of truth. Yet that relation is, I believe, more opaque than any other in the whole domain of semantics. In fact, avoiding the perplexities that it raises primarily motivates our settling for a theory of truth whose domain is sentences in the first place.

B. Tarski's Theoretical Definition of Truth

To return to our central concern, I am suggesting that in following Benacerraf we require of a satisfactory philosophy of mathematics a definition of truth of the (nonmodal) form indicated, which is theoretical in the sense I have explained, though it be neither a priori, necessary, nor analytic. Just this much would not seem to satisfy Benacerraf though, for we observed him to require that "the theory of mathematical truth be in conformity with a general theory of truth--a theory of truth theories, if you like" (p. 666). Benacerraf seems to be worried that the sentences held to be mathematically true on the basis of the proffered account will turn out not really to be true at all. I am baffled as to what Benacerraf has in mind and so am uncertain how to satisfy him on this score. I certainly hope that whatever it is that he is looking for, Benacerraf would be satisfied, with regard to his semantical requirement, by a theory of truth of the kind we have envisioned for a larger language that contained the language of mathematics as a proper part. Given a way of distinguishing the sentences of mathematics from the other sentences, say by their terminology, the account of mathematical truth could simply have it that a sentence was mathematically true just in case it was true, and, what is more, was a sentence of mathematics.[28] Presumably this would insure that truth in mathematics is like truth elsewhere. Perhaps we should worry about how large the larger language would have to be; even if its expressive power went beyond mathematics, it might not go far enough beyond to quell Benacerraf's worries. Some might wish that it be so large as to express every proposition whatever. If so, they will be disappointed, for I am certain, though I will not argue for it here, that in no language can we say everything there is to be said.

Perhaps Benacerraf recognizes this and would instead seek a theory of truth that was general in that it ranged over the sentences of all

[28] There is a certain impropriety in speaking of the language of mathematics where what is meant is some proper part of a larger language. There is such thing as mathematical language, in that there are languages--formal languages--all of whose well-formed sentences express propositions purely of mathematics and all of whose terms are recognizably mathematical terms. Yet obviously any subset of the grammatical sentences of English--purporting to be the mathematical fragment--either will include sentences some of whose truth-functional components have no mathematical terminology or will exclude sentences some of whose truth-functional components have exclusively mathematical terminology. Be this as it may, it poses no more of a problem for the proffered account of mathematical truth than it does for any other.

languages or, if it be possible for one sentence to occur in more than one language and differ in truth value over the different occurrences, a theory of truth that was general though relative in that it related all sentences to all their containing languages in respect of truth. I believe and have argued elsewhere that there is no reason to expect that a general theory of truth of either form can be developed.[29] To consider that issue in more detail would take us too far afield.

Perhaps a more modest achievement would satisfy Benacerraf. This would be a theory of truth for a language confined to expressing mathematical propositions, say the language of Principia Mathematica or a language for Peano arithmetic, which was structurally similar to theories of truth for other languages, though these languages be similarly confined in their expressive power. Perhaps we should require that some of these other languages have the power to express nonmathematical propositions, for example, propositions of physics or of economics. If this much would satisfy Benacerraf's requirement for a semantics for mathematics, it has, I believe, quite ably been fulfilled by Tarski's theory of truth. (One qualification suggests itself. I do not know of a specific example of a theory of truth for a nonmathematical language that is structurally similar to Tarski's theory, that is, I cannot cite one in the philosophical literature. Yet I have no doubt that one could easily be constructed.)

Besides working in parallel ways for both mathematical and nonmathematical languages, the Tarski definition has another virtue, which directly answers Benacerraf's concern that the concept being defined turn out

[29] In "Field's Theory of Truth." See pp. 70-76, above.

to be genuine truth. As might be expected, what I here refer to is its fulfillment of what Tarski called the condition of material adequacy. The Tarski definition, together with supporting syntactical and logical propositions, entails all instances of Tarski's so-called Convention T. These are the disquotational paradigms exemplified by the now notorious case,

'Snow is white' is true if and only if snow is white.

We have already seen that intuitively obvious consequences certify a definition as a genuinely theoretical claim, giving it the kind of epistemological support that is characteristic of theoretical claims. But these consequences do more. They articulate what is most central in our ordinary understanding of truth, namely, that for any sentence to be true, the state of affairs it describes must actually prevail. Unlike the latter articulation, which depends on dubiously metaphorical (or tendentiously metaphysical) phrasing, these consequences articulate that central idea in a metaphysically innocuous way. This theoretical structure seems utterly impossible unless what is being captured by the definition is indeed truth. And Benacerraf's worries are thereby laid to rest.

Of course Benacerraf is aware of Tarski's theory. Having set out his semantical condition, he says, "I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language" (p. 667). But then he goes on to say that he believes that fulfilling the semantical requirement in this way "is inconsistent with all the accounts that I [Benacerraf] have termed 'combinatorial'" (p. 667). Why is this so? Benacerraf's argument for it fails completely to convince me. He mounts an argument only in his closing paragraphs and directs it only to a particular combinatorial view, namely, conventionalism. Even this narrow argument is difficult to follow.

Rather than lingering over Benacerraf's argument that Tarski's theory is incompatible with combinatorial views, I have chosen to address this issue conversely. Benacerraf has deemed the standard semantics, by which he means something at least structurally similar to the Tarski theory, to be the friend of the standard, or Platonistic, philosophy of mathematics. And he has insisted that anti-Platonistic philosophies of mathematics are incompatible with the standard semantics. The question naturally suggests itself, does Tarski's theory of truth entail, or more weakly, lend evidential support to Platonistic views in the philosophy of mathematics? It is that question to which we direct our attention in the next section.

VII. PLATONISM AND TRUTH

It is commonly believed that an acceptance of Tarski's theory of truth for mathematical languages carries with it an acceptance of mathematical Platonism. Mark Steiner opens his discussion of Platonism with the following representative remarks:

> According to ontological Platonism, the truths of mathematics describe infinitely many real mathematical objects. Since the number of material objects may very well be finite, most mathematical objects could not be material . . . Furthermore, the only way that mathematical statements could be true is by describing such mathematical objects (see Tarski, "The Concept of Truth in Formalized Languages," . . .). It follows, then, that whether (ontological) Platonism is tenable is the same question as whether the axioms of mathematics are _true_. This conclusion puts ontological Platonism in a very favorable light.[30]

Evidently Steiner believes that under Tarski's definition of truth, the axioms of mathematics can turn out to be true only under a Platonistic conception of those axioms. In other words, given the Tarski account of truth, we reject Platonism only at pain of denying truth to even the most fundamental and evident mathematical statements.

It is the burden of this section to determine whether this common belief, that Platonism is implicated in Tarski's theory, is well-founded. Of course doing so requires us to get clear about what philosophical doctrines comprise Platonism. I have found several different doctrines put forward in the name of Platonism, some superficial and others of subtlety and interest. The superficial doctrines can be quickly set aside; there is not even _prima facie_ reason to infer them from a theory of truth. The Platonistic principles that I concentrate on include the principle of the equivalence of significance and grammaticality, and certain principles associated with philosophical realism: the Principle of Bivalence, the principle that truth is independent of verifiability, and the principle that existential commitment in mathematics cannot be explained away. These all have the advantage of precision and definiteness. Though in

[30] Steiner, _Mathematical Knowledge_, p. 109.

each case it is tempting to infer the principle in question from the Tarski theory, in the end its very precision enables us to reconcile its demands with the Tarski theory.

Steiner uses the term "ontological Platonism" to distinguish the doctrine there characterized from the apparently independent doctrine that we know the propositions of mathematics, or at least the fundamental ones, through a special kind of intuition, analogous to sensory perception of visible and tangible objects. The latter doctrine is often attributed to Gödel and is clearly the heir of Plato's views about our knowledge of the Forms. In keeping with our avoidance of epistemological issues, we will follow Steiner in separating this epistemological doctrine out. Because I am not sure that what remains is an exclusively ontological doctrine, I will not adopt Steiner's qualifier. What, for our purposes, Platonism consists in remains to be explained, but we have determined that we do not mean an epistemological doctrine.[31]

What then is the doctrine that is supposed to follow from Tarski's theory of truth? Steiner's characterization, that mathematical statements "describe infinitely many real mathematical objects" clearly needs some elaboration. What he seems to emphasize, that there are infinitely many things in the domain of mathematics, is no consequence of Tarski's truth definition. It seems to me not to be, properly speaking, a part of

[31] I take it that no one is tempted to believe that the epistemological doctrine is a consequence of Tarski's theory of truth, unless it be by way of some other Platonistic doctrines. Both Steiner and Benacerraf do strongly suggest that epistemological Platonism is a pretty direct philosophical consequence, though not a logical consequence, of the doctrine here to be considered, whence Benacerraf's dilemma.

Platonism anyway, because whether there are infinitely many numbers is not a point at issue in the philosophy of mathematics. Everyone agrees that if there be any numbers at all, there are infinitely many, though perhaps that infinity is never completed. Perhaps, that is, there is no infinite set, containing all the numbers.

One popular conception of Platonism focuses on the abstract character of mathematics. Philip Kitcher says, "Arithmetical Platonism is the thesis that statements of arithmetic are true or false in virtue of the properties of objects, the natural numbers, which do not exist in space-time, and which therefore deserve to be called 'abstract objects'."[32] It is this aspect of Platonism that is the point of Steiner's argument from the finitude of material bodies. There are much more convincing arguments that numbers and functions and such are abstract, arguments that depend on no such questionable assumptions. After all, numbers have no physical properties; they weigh nothing, have no charge or mass; they have neither spatial location nor spatial extension. Though they may quite properly be said to exist now, if at all, and indeed to exist always, they are timeless in that they neither come nor go, and, with regard at least to those properties treated in mathematical theories, they change not at all. All of this marks the numbers as abstract. And I assume that no one would be tempted to deny them these marks of the abstract. Nevertheless, nothing in a Tarski theory of truth for any mathematical language entails any of this. It says nothing of abstraction or of weightlessness or of timelessness.

[32] Philip Kitcher, "The Plight of the Platonist," Nous, 12 (1978), 119-136; see p. 119.

So far remarkably little seems to be at stake in the acceptance of Platonism. In search of a more robust version, we turn to a characterization by Charles Parsons:

> Platonism is the methodological position which goes with philosophical realism regarding the objects mathematics deals with. Mathematical objects are treated . . . as if the facts concerning them did not involve a relation to the mind or depend in any way on the possibilities of verification. . . . This is taken to mean that certain <u>totalities</u> of mathematical objects are well defined, in the sense that propositions defined by quantification over them have definite truth-values. Thus, there is a direct connection between platonism and the law of excluded middle.[33]

Whatever we might say about the structure of Parsons' remarks, he certainly mentions some interesting matters.

A. Three Aspects of Platonism

Picking up cues from Parsons, I regard Platonism as the combination of three doctrines regarding the nature of mathematical theory and its proper development, not only with respect to ordinary arithmetic, but more particularly with respect to set theory, transfinite arithmetic, and analysis--all areas in which the notion of infinity plays a central role.

(1) Platonism holds that well-formed sentences in mathematical languages, including infinitary ones, have truth values; some of them are true and some of them are false. In another way of putting this, all well-formed mathematical sentences express propositions.

[33] Charles Parsons, "Foundations of Mathematics," in <u>The Encyclopedia of Philosophy</u>, ed. Paul Edwards (New York: Macmillan, 1967), v. 5, p. 201.

(2) Platonism defends philosophical realism with respect to mathematical objects.

(3) As a methodological position, Platonism endorses certain forms of mathematical reasoning that are not universally accepted.

With regard to the first of these doctrines, we must ask whether the equivalence of significance and grammaticality is forced upon us, or even lent plausibility, by definitions of truth of the sort proposed by Tarski. With regard to the second, we must ask what specific principles are included in philosophical realism, and then ask of each of them the same questions about its relation to Tarski's theory.

With regard to the third doctrine, Platonism endorses in particular those forms of reasoning whose appearance of soundness depends on the supposition that infinite collections of mathematical objects stand complete at the moment of their specification, rather than merely representing a process of collecting or constructing, each step of which, as it were, consumes a definite period of time, and which will not be completed in any finite number of steps, and thus will at no time be completed. Two central examples are Platonism's acceptance of non-constructive existence proofs and its acceptance of impredicative definitions. No one unfamiliar with the disputes between Kronecker and Hilbert or between Poincare and Russell near the turn of this century can fully understand that, far from constituting arid metaphysical debates, methodological disagreement over these forms of reasoning affected the practice of mathematicians in areas of widest research interest. Nevertheless, exploring this third aspect of Platonism goes beyond the scope of this study. The mathematical issues are both complex and abstruse. And moreover, given that these are issues about the legitimate forms of definition and inference, there is little reason on the face of it to expect them to be prejudiced by a theory of truth like Tarski's.

B. Significance and Grammaticality

Is it necessary in any truth definition modeled after the one given by Tarski for the language of the calculus of classes that all of the well-formed sentences in the language for which truth is being defined be given truth values? It was assumed by Tarski that the language for which he defined truth had a peculiar characteristic which ceases to be obvious as soon as we call it into question. Under Tarski's definition, expressions qualify for the assignment of truth values purely on the basis of lexicon and syntax. The definition applies to all sentences in the language; as is usual in formal languages, any expression constructed out of a certified vocabulary according to specified recursive rules of sentence formation qualifies as a sentence in Tarski's language. In effect Tarski's definition takes it for granted that every well-formed sentence in his language says something definite, expresses a definite proposition. Extrapolating from Tarski's assumption, we get the general principle that any expression that results from the rules of sentence formation for its language is fully significant and usable for asserting, whether truly or falsely. A succinct way of putting this is to say that significance goes with grammaticality.

I deny that it is obvious on a second look that significance accompanies grammaticality for the (perhaps unconvincing) reason that many philosophers have rejected this correlation of the two notions. They have had various motivations. Perhaps the most well-known example is that of the logical positivists with their principle that only sentences that are either empirically verifiable or analytic are significant. According to the principle of verifiability, a sentence that in principle cannot be shown to be true or be shown to be false is meaningless, though it may be indistinguishable in point of grammar from fully meaningful sentences. When someone utters such a sentence, they do not succeed in asserting anything, and so the question of truth or falsity cannot even come up.

Despite his familiarity with both the principle of verifiability and its defenders, Tarski ignored this issue, perhaps because he was defining truth for a mathematical language, for which the issue of empirical significance did not arise. Or perhaps he was out of sympathy with the doctrine of verificationism and with the search for philosophical criteria of significance. In any case I know of no explicit articulation of the view that significance is coextensive with grammaticality before Quine's article, "The Problem of Meaning in Linguistics."[34] There Quine rejects the relational notion of having a meaning, in favor of the ontologically uncommitted notion of significance, and sketches a behavioral criterion for the latter. The resulting notion is (dimly) recognizable as the familiar notion of grammaticality. But Quine says nothing in defense of this

[34] Willard Van Orman Quine, "The Problem of Meaning in Linguistics," in his From a Logical Point of View, 2nd ed., rev. (Cambridge: Harvard University Press, 1980), pp. 47-64. See Sections 2 and 3 of that paper, particularly pp. 51-52. Incidentally, the paper grew out of a lecture given in 1951.

criterion of significance against philosophical challenge; he recommends it purely in virtue of its behavioral basis.

Presumably our readiness to accept grammaticality as a sufficient condition of significance rests on our Fregean preconceptions about semantic compositionality. Meaningful parts composed according to sanctified rules yield meaningful wholes. Even so, we are prepared to admit that in those languages with indexical terms, fully grammatical sentences might not be assertorically usable in certain contexts. Here opens a cleavage between being significant in the sense of being fully meaningful and being significant in the sense of being fully interpreted or of saying something definite or of expressing a proposition (pick your favorite). Tarski can justifiably ignore that cleavage because his language pointedly excludes indexical and other context-dependent elements.[35] But can a Tarski theory accommodate a philosophically motivated challenge to the Fregean preconceptions. Suppose a philosopher of science embraces a principle of empirical verifiability. Or suppose a mathematician develops a new language and accords significance only to sentences of the language that are decided by a favored set of axioms, regarding the others, however grammatical, as not fully interpreted. Or suppose a philosopher of mathematics distinguishes the sentences of classical mathematics that can be checked by finitary arithmetical methods from the others, arguing that the others are devoid of literal sense or merely ideal. In adopting

[35] Compare "Field's Theory of Truth," pp. 54-56, above.

any of these views have we implicity rejected theories of truth of the Tarski type?

The correct response seems clearly to be no. Providing a criterion of significance is no job for a theory of truth, nor is embracing a criterion that has been incidentally provided. Tarski truth definitions typically have the form,

For any sentence s of the language L,
s is true if and only if $\phi(s)$.

The qualification attached to the initial generalizing phrase serves to express the criterion of sentencehood and the criterion of significance at once; any expression that qualifies as a sentence can try to measure up to the condition necessary for truth expressed by '$\phi(s)$'. Yet however characteristic of Tarski truth definitions this qualifier is, it is incidental to the central idea behind them. A philosophy of mathematics that denied the sufficiency of grammaticality for significance could merely elaborate the restriction on the quantifier, making it of the form,

For any sentence s of the language L such that $\alpha(s)$,

where '$\alpha(s)$' would be replaced by the favored criterion of significance. So long as that criterion did not smuggle an alternative definition of truth, it seems clear that the modified definition would still imitate the crucial features of the Tarski paradigm and would thereby qualify as a Tarski theory of truth.

Therefore at least the first aspect of Platonism that we have isolated does not follow from Tarski's theory of truth. An opponent of

Platonism could supply a definition of truth that was structurally similar to Tarski's definition, one that defined truth in terms of satisfaction, indeed one that imitated Tarski's in its specification of satisfaction conditions for the atomic sentences of the language, and yet the definition could differ in this incidental way of restricting its field of application to those sentences that were significant according to whatever criterion of significance the anti-Platonist proposed. His definition would entail not all instances of the disquotation schema, but rather only those instances whose sentential referents survive the test of significance; happily these would be exactly those instances that were significant themselves. Of course the definition would not have the consequence that all well-formed sentences of the mathematical language have truth values. And thus Tarski's theory lends no plausibility to this aspect of Platonism, however plausible it may be in itself.

C. Realism with Respect to Mathematical Objects

The second aspect of Platonism that we registered was its defense of philosophical realism concerning mathematics. The term that is perhaps most frequently employed in explanations of Platonism is the adjective, 'real'. We saw Steiner speak of mathematics describing "infinitely many real mathematical objects."[36] The unrelenting presence of this adjective inevitably evokes the remarks of Austin in addressing an altogether different philosophical topic. Austin observed that "with 'real' . . . it

[36] Steiner, Mathematical Knowledge, p. 109.

is the negative use that wears the trousers. That is, a definite sense attaches to the assertion that something is real. . . only in the light of a specific way in which it might be, or might have been, not real."[37] (Are we to suppose that the numbers, not being real, are complex?) Having long been sympathetic to his observation, I can hardly let us rest content with this kind of account of mathematical realism.

An analogy comes to mind, one drawn in many discussions which intend to flesh out the doctrine of realism with respect to mathematics. Dummett especially likes to compare the realist's conception of mathematics to the ordinary conception of astronomy. On this conception there is a realm of numbers and other mathematical entities, like there is a universe occupied by galaxies and stars. The mathematician directs his gaze at these mathematical quasars through the telescope of mathematical intuition, making discoveries about their nature and relationships. As Dummett puts it, "Mathematical structures, like galaxies, exist, independently of us, in a realm of reality which we do not inhabit but which those of us who have the skill are capable of observing and reporting on."[38] Drawing a somewhat similar analogy, Wittgenstein spoke (in derision) of "arithmetic as the natural history, or minerology, of numbers."[39]

[37] J. L. Austin, Sense and Sensibilia (Oxford: Oxford University Press, 1962), p. 70.

[38] Michael Dummett, "The Philosophical Basis of Intuitionistic Logic," in his Truth and Other Enigmas, p. 229.

[39] Ludwig Wittgenstein, Remarks on the Foundation of Mathematics (1956; rpt. Cambridge: The M.I.T. Press, 1967), p. 116. Quoted passage is from Part III, Section 11.

Dummett acknowledges that all we have in the astronomical analogy is a metaphor or a picture.[40] Wittgenstein also speaks of his analogy as a picture, for him one to be excoriated. Indeed everything that the analogy literally gives us, that is, every way in which mathematics is like astronomy reduces to saying that the favored mathematical propositions are true. Unfortunately, whatever the nature of truth, this last is not a proprietary doctrine of realism. Steiner is right in seeing our adherence to this doctrine, at least as regards elementary arithmetical propositions, as inexorable. And many constructivist opponents of Platonism accord truth even to certain infinitary propositions, namely those that are demonstrable by their own austere standards.

In response several worthy attempts have recently been made to state precisely wherein realism consists.[41] These attempts converge on two tenets that are peculiar to and characteristic of philosophical realism both in the philosophy of mathematics and elsewhere. The first tenet is what Dummett calls the Principle of Bivalence, the semantic principle that

[40] He goes on to ask how to give substance to this metaphor (p. 231ff). Though his discussion is of interest, it would take us in a path tangential to our own, which circles around Benacerraf's worries.

[41] This has been a favorite topic of Dummett's; it explicity emerges in a number of the essays in Truth and Other Enigmas and lies just beneath the surface in almost all of the others. Hilary Putnam's essay, "What Is 'Realism'?" was delivered before the Aristotelian Society in 1976 and became Lecture II and half of Lecture III of his John Locke Lectures; see Hilary Putnam, Meaning and the Moral Sciences (London: Routledge, 1978). I also recommend the opening chapter of Crispin Wright, Wittgenstein on the Foundations of Mathematics (Cambridge: Harvard University Press, 1980).

every proposition is either true or false.[42] The second tenet, more difficult to state and accordingly less uniformly formulated among proponents of realism, is some principle maintaining the independence of truth from verifiability or demonstrability, even highly idealized versions of verifiability. These two tenets have recently been taken to be the literal core of philosophical realism in all its philosophical domains. They both have the virtues of being precisely expressible and easily compared with the Tarski theory. It must be admitted, however, that these two tenets disregard a philosophical move that was more widely attempted earlier in this century than it has been in the past decade. I am speaking of the attempt to explain away mathematical objects, to treat them as fictions, by giving analyses of the propositions of mathematics, in which there is no existential quantification over anything corresponding to the numbers or to other mathematical entities. The denial of this possibility must be acknowledged as another important tenet of realism. Realism maintains, in short, that numbers are real, not fictions of convenience. For our purposes the question of the evidentiary relation of this aspect of Platonism--its defense of philosophical realism--to Tarski's account of truth reduces to the question whether these three principles--the Principle of Bivalence, the independence of truth from verifiability, and the impossibility of explaining away mathematical objects--are either logical or philosophical consequences of definitions of truth of the Tarski type. Let us take up the questions in order.

[42] He reserves the title "Law of Excluded Middle" for the schematic principle, "P or not-P". See Dummett, Truth and Other Enigmas, p. xix.

D. The Principle of Bivalence

We have seen that Benacerraf uses Gödel as the archetype of his Platonist philosopher of mathematics; he is supposed to have been driven by his standard convictions about semantics and by his desire for an harmoniously related epistemology, to a far-fetched epistemological doctrine, in fact to the doctrine that Steiner calls epistemological Platonism. This portrayal of Gödel is somewhat surprising in that Gödel says so very little about semantical issues, or more particularly about the nature of mathematical truth, in the essay under consideration. There does appear an interesting discussion of the meaning of the term 'set'.[43] Gödel takes seriously, if only for long enough to argue against it, the supposition that the concept of set, and accordingly the meaning of the term 'set', has not been fully determined by the practice of set theoreticians. This discussion takes note of the existence of alternative models of the accepted set-theoretic axioms and thereby implicitly employs notions of referential semantics. But there is very little attention paid to the nature of those notions, or even to the question of how to determine which of the models actually is the domain of the quantifiers in set theory. What he does do, and this is the central burden of the essay, is to insist that Cantor's continuum problem is a legitimate mathematical problem that he expects to be solved by ordinary mathematical methods, though not to be decided by the currently accepted axioms. (Of course this

[43] Gödel, "Cantor's Continuum Problem," p. 262 and pp. 265-266.

last is now a standard result of set theory.) He gives recognizably mathematical reasons for doubting the truth of the continuum hypothesis, and then suggests that it will be shown to be false by the discovery of new axioms, which once formulated will "force themselves upon us"[44] in exactly the way the usual axioms do.

Gödel's only philosophical defense of his belief that the truth of the continuum hypothesis constitutes an adequately formulated and legitimate mathematical problem lies in the following passage.

> It is to be noted, however, that on the basis of the point of view here adopted, a proof of the undecidability of Cantor's conjecture from the accepted axioms of set theory . . . would by no means solve the problem. For if the meanings of the primitive terms of set theory . . . are accepted as sound, it follows that the set-theoretical concepts and theorems describe some well-determined reality, in which Cantor's conjecture must be either true or false. Hence its undecidability from the axioms being assumed today can only mean that these axioms do not contain a complete description of that reality.[45]

Abstracted from its epistemological content then what Gödel's Platonism, as expressed in this essay, consists in is simply an adherence to the Principle of Bivalence, along with the implicit use of our earlier principle that grammaticality insures significance. What Gödel must be urging is that once any doubts about the meanings of the primitive terms of set theory have been dispelled, then well-formed set-theoretic sentences,

[44] Compare Gödel, "Cantor's Continuum Problem," p. 271.

[45] Gödel, "Cantor's Continuum Problem," pp. 263-264.

like that articulating Cantor's continuum hypothesis, are guaranteed of being either true or false. For the continuum hypothesis to "describe [or misdescribe] some well-determined reality" seems simply to be for it to be either true or false, despite our inability to determine which on the basis of current mathematical understanding. And what could Gödel's confidence about this particular case rest upon, other than a conviction about the general principle that every proposition is either true or false? This of course is the Principle of Bivalence.

Of all of the aspects of Platonism that we have isolated, the Principle of Bivalence is the one which we are most tempted to conclude from Tarski's theory of truth. In fact, it (or a very near relative) is a proud result in Tarski's original paper:

> Theorem 2 (The principle of excluded middle). For all sentences x, either $x \in Tr$ or $\bar{x} \in Tr$.[46]

Since a sentence is false on Tarski's definition just in case it is not true, this theorem is not quite Bivalence as we have stated it. But on Tarski's definition the negation of a sentence is true if and only if the sentence is not true, and therefore the Principle of Bivalence is an immediate consequence of Tarski's Theorem 2. Here, if anywhere, we seem to have a serious philosophical consequence of Tarski's theory. In fact this seems to be exactly the kind of general consequence that we were looking for earlier (p. 127, above) to qualify a truth definition as a

[46] Tarski, "The Concept of Truth," p. 197.

theoretical definition. The Principle of Bivalence is a generalization of very great interest that certainly has not been justified to the satisfaction of everyone. In fact what is probably the central controversy in the philosophy of mathematics turns upon it. Unfortunately this consequence turns to ashes in our mouths, for when we examine Tarski's derivation of the Principle of Bivalence, we find that it depends on an assumption that is equally controversial and that is explicitly rejected by many opponents of Platonism.

Specifically, Tarski's derivation of Theorem 2 depends on drawing the inference that if not every infinite sequence satisfies a particular sentence, then there is some infinite sequence that does not satisfy it. Of course this inference is an instance of one of the central implications of classical quantification theory, that '$\sim(\forall x)Fx$' implies '$(\exists x)\sim Fx$'. However central, this implication is perhaps the most disputed result of classical quantification theory, one that is specifically denied by typical opponents of Platonism. The cases in which anti-Platonists deny this implication are in fact the cases where the quantifiers in question range over an infinite domain. And the case in point is one such, because any Tarski truth definition where either the language under consideration has infinitely many variables or its quantifiers range over infinitely many objects, requires infinitely many satisfaction sequences.

Speaking more formally, the derivation of the needed quantificational implication inevitably employs principles of inference which equally yield the Law of Excluded Middle properly speaking, that is to say, the logical schema, 'P or not-P'. It comes as no surprise that the Tarski definition of truth provides a proof of the Principle of Bivalence only upon the employment of rules of inference which immediately yield the corresponding logical principle, Excluded Middle. (The rule of inference in question, in most formalizations, is that licensing the inference of a statement from its double negation.) Most anti-Platonists have independent reasons for objecting to those rules of inference. Disallowing inferences employing those rules has the perhaps unintended consequence of rendering the Tarski definition of truth powerless to produce the Principle of Bivalence and thus rendering it perfectly acceptable to anti-Platonist opponents of Bivalence.

E. The Independence of Truth from Verifiability

A second tenet of realism that has been emphasized in recent treatments is the principle that truth is independent of verifiability. Indeed, it is related to the earlier tenet, the Principle of Bivalence, in that if not every proposition could be either verified or falsified even in the fullness of time, the Principle of Bivalence would be lost. This connection has led some actually to identify these two tenets, to presume that the independence of truth from verifiability could not be stated other than as the Principle of Bivalence.

One philosopher who has found an interesting articulation of the idea that truth is independent of any epistemological properties is Hilary Putnam. Indeed the proposition that he puts forward is offered in answer to the question, "What is realism?" His principle is,

> A statement can be false even though it follows from our theory or from our theory plus the set of true observation sentences.[47]

[47] Putnam, Meaning and the Moral Sciences, p. 34.

Here again we have given quite definite content to one of the root ideas informing philosophical realism.

In surveying a Tarski definition of truth, one might be tempted to infer from the theory that truth is independent of verifiability, because neither verification nor any other epistemological notions (for example, provability, derivability from favored axioms, confirmation) appear prominently in the definition. (Of course these notions would appear in that part of the definition intended to determine the truth values of statements, if such there be in the language for which truth is being defined, that themselves incorporate epistemological terms.) The truth conditions consist of seemingly objective facts, whatever facts are describable in the language under consideration.

Yet the absence of epistemological terms in the theory of truth does not itself guarantee that truth is independent of epistemological notions, that is to say, it does not rule out the possibility that in fact all and only verifiable propositions are true. Moreover this might turn out not to be a mere fact, as it were accidental, but a deep philosophical result. The best way of seeing this point is to consider a mathematical language in which a complete (that is, decidable) mathematical theory could be stated. Suppose also that the axioms for this complete theory could somehow be shown to be self-evident. I have no idea how this might go, but I am just supposing that some philosophical argument for or explanation of the self-evidence of the axioms could be given. Think for example of Kant's account of the _a priori_ character of geometrical statements. (Of course I am not endorsing Kant's view.) In this case truth would _not_ be independent of verification. In fact all of the sentences of this mathematical language would have completely convincing _a priori_ proofs or disproofs. Nevertheless, a Tarski style theory of truth would remain in place for this language with exactly the same content as it would have if these happy epistemological circumstances did not obtain.

There is no reason, at least none captured by the Tarski theory, why this situation might not hold for all of mathematics. Of course in the face of Gödel's incompleteness proof, either the Principle of Bivalence or some inferential principles employed in that proof would have to be given up. But we have already seen that these are not guaranteed by Tarski's theory either.

It might be objected that in any case the Tarski theory shows that it is no part of the meaning of truth that truths are demonstrable, that is to say, that truth is at least not analytically reducible to verifiability. The reply is that we have made no claim that the Tarski theory gives analytically necessary and sufficient conditions for truth, thereby ruling out other analytically necessary and sufficient conditions. What distinguishes the definition, on my view, from other sets of truth conditions is not that it is analytical, but rather than it is theoretical.

On a second look at the formulation of this realist tenet taken from Putnam, we lose _all_ confidence that the independence of truth from verifiability is a consequence of Tarski's theory, for Tarski says nothing in particular about the contents of "our theory" and the bearings of its appearing in our best theory on the truth or falsity of a statment. It is for precisely this reason that Putnam was pushed to the quoted principle in formulation of philosophical realism. A few sentences earlier he says, "Nor is [understanding truth and the logical connectives realistically] just a question of accepting Criterion T . . . or a question of accepting a Tarski-style truth definition for one's language."[48]

E. Explaining Away Mathematical Objects

One important way of denying the reality of some kind of objects is by showing how we can say everything that we want to say without quantifying over those objects. We can even accord truth values to those sentences that have the appearance of quantifying over the unwanted objects, so long as for each of them we give an analysis in which no existential quantification occurs that requires these objects as values for its variables. Since the sentence resulting can be true, though there exist none of the problematic things, the original can be said to be true as well. Since the official version is put forward as an analysis of the original, the original can inherit its truth value from the official version. (Of course this requires a more robust conception of analysis than Quine's notion of regimentation provides, but no matter.) Here is a way of having our cake and eating it too. We stick with the standard way of talking in any but ontologically delicate contexts, secure that our utterances are fully significant and truth-value laden. But when pressed regarding the ontological force of our existential quantifications, we fall back on the official versions of the sentences we uttered and point out that there no worrisome quantification can be found. In this manner some philosophers explain away the numbers or classes or whatever mathematical entities they wish to condemn.

[48] Putnam, Meaning and the Moral Sciences, p. 34.

In the now familiar pattern, this kind of anti-realist doctrine with respect to numbers or other mathematical entities, might be thought to be ruled out by a Tarski-style theory of truth. If we explain away the numbers, then ordinary arithmetical sentences, with their apparent quantification over numbers, inherit their truth values from official equivalents which do not quantify over the numbers. A Tarski theory says nothing of this, but blandly assigns truth values by way of the surface syntax of the ordinary arithmetical sentences, using the clause for existential quantifiers in the usual way. Hence, the Tarski theory would seem to be incompatible with the doctrine that they inherit their truth values from the ontologically innocent official equivalents, and so would seem to entail the Platonistic principle that existential commitment in mathematics cannot be explained away.

Again a closer look reveals no such consequence of the Tarski theory. As we have seen before, the theoretically central truth conditions articulated by the Tarski theory may coexist with other sets of truth conditions discovered by attending to different issues. In this case a concern with ontology leads to that truth condition which we described in terms of inheritance from the official analysis. So long as that truth condition yields the same result as the Tarski truth condition, that is, the same truth value, then they can both hold good. And it must yield the same result or else we would not accept the analysis as adequate. Of course the ontologically niggardly philosopher who proposed the analysis in the first place might well object to the Tarski theory on the grounds of its own excessive ontology. In response we could invite him to apply his own methods to the language in which the Tarski theory is expressed.

According to that theory, the condition of the truth of the existentially quantified sentence, 'There is a prime between 5 and 11', is that there exist some sequence of numbers which satisfies the open sentence, 'x is a prime between 5 and 11'. Apparently, for that condition to hold, there must be sequences of numbers included in the range of the existential quantification, 'there exist some sequence of numbers', one sentence back. But the Tarski truth condition that speaks of sequences of numbers is not at all unusual in that respect, mathematically speaking. Lots of mathematical theories speak of sequences of numbers, and if the proposed method of explaining away the objects of mathematics is any good, it can be expected to handle such mathematical theories. It can proceed to handle the Tarski theory in exactly the same way.

Biographical Note

Douglas Cannon was born in Yakima, Washington, in 1946, the son of Rowland and Elithe Fillmore Cannon. He attended public schools in the Intermountain West and graduated from Idaho Falls High School in 1964. He attended the University of Utah in the late 1960's and graduated from Harvard College in 1973. During those years he worked on the development of computer systems in several settings, among them the Utah-Idaho Sugar Company and the University of California at Santa Cruz. He undertook graduate studies in philosophy at M.I.T. in 1973, serving during his first year as Research Assistant in the Artificial Intelligence Laboratory. From 1978 to 1980 he was an Associate with Index Systems, a management consulting firm in Cambridge, Massachusetts. Since 1980 he has served as Assistant Professor at the University of Puget Sound in Tacoma, Washington. He is married to Mary Ellen Sullivan of Santa Rosa, California. They have two children, Mary Eleanor and Sarah Elithe Cannon.